Self tuning control applied to heating systems.

Murtagh, K. J.

Award date:
1985

Awarding institution:
University of Bath

SELF TUNING CONTROL APPLIED TO HEATING SYSTEMS

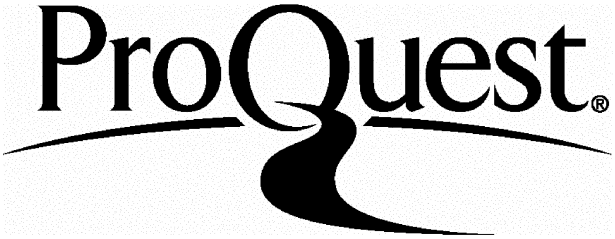
submitted by K. J. Murtagh

for the degree of Ph.D

of the University of Bath

1985

K. Murtagh.



To Romola

Acknowledgements

The author gratefully acknowledges the help and guidance of the staff of the School of Electrical Engineering in the completion of this research project. Special thanks are given to Dr R T Lipczynski for his help in supervising the project and the completion of the thesis.

The technical advice and financial support of Satchwell Control Systems was instrumental in the success of this project and this is gratefully acknowledged.

CONTENTS

ABSTRACT

The use of self tuning control to aid the commissioning and to improve the performance of environmental heating controls is investigated. The theory of self tuning control is well developed, this study considers its application to air conditioning plant, and in particular the measures necessary to devise a controller that is sufficiently robust to be competitive with conventional controllers. Nonlinearities peculiar to air conditioning plant are considered in detail and the effect on the self tuning control is investigated. The commissioning of heating controls can be simplified by using self tuning control, as its predictive nature compensates for the large phase lag present in most heating plant. Commissioning of controls is described in detail, and a novel commissioning procedure is presented for use with self tuning control. An important feature of heating plant is the air mixing process. A test facility that exhibits the major characterics of a practical air mixing process, such as an occupied air space (room) is constructed and its thermal characteristics modelled. Modelling of practical heating plant is described and techniques based on crosscorrelation and least squares methods are presented as a means of obtaining optimal estimates of the model parameters. The results of parameter identification of the test facility highlight several of the problems of applying self tuning control in practice, and in particular to heating plant.

The self tuning control presented is shown to have a less sensitive commissioning procedure and better, long term performance than conventional PI control both in simulation and in real time control of the test facility.

1 Introduction

In recent years a growing awareness of the Earth's diminishing fossil fuel deposits has stimulated research into finding new energy sources, and developing techniques for using existing resources more efficiently. Of paramount importance to the latter field of research is the advance of computer technology in particular the development of the microprocessor. The use of low cost microprocessor based equipment, has allowed considerable computational power to be brought to bear on energy reduction problems that previously required the use of main frame or mini-computers, the cost of which being far greater than any saving that could be made by their use.

In the UK and much of Europe, a large proportion of a nation's expenditure on energy is for space (room) heating alone, hence much effort has been directed to making space heating systems more energy efficient, whilst retaining or improving the comfort conditions of the occupants. The first step to reduce heat energy loss is to install sufficient thermal insulation throughout the building structure. A further step is to orientate new buildings to make best use of solar radiation and/or supplement the heating system output by heat energy obtained from solar collectors (67).

Given that the building is adequately insulated, and the best use of supplementary heating is made, then the next

step in reducing energy loss is to operate the heating system no earlier than is necessary to guarantee that the desired temperature is obtained within the space at the point of occupancy. The time taken (pre heat time) for the space temperature to reach the desired occupancy value is a function not only of the heating system, but also of the building structure and the outside and inside environmental conditions. Thus to calculate the preheat time the thermal characteristic of the building and the heating system must be known. To avoid the manual calculation or measurement of thermal characteristics, self adaptive preheat strategies have been devised to optimise plant operating periods (26, 25, 78).

It has also been established (50, 9) that a reduction in energy can be obtained, whilst comfort conditions remain acceptable, by the more efficient operation of temperature regulation controls. This is admirably demonstrated by the results of a recent study (72) of the energy used by thermostatically controlled domestic heating systems. The object of the study was to observe how the energy used is related to the thermostat differential. The thermostat differential is the temperature difference between that which would cause the heating system to switch from on to off, or off to on.

The results show that in households where a large differential was used the thermostat was set higher than for households where a small differential was used , this being due to the reaction of the domestic user to the

minimum perceived temperature rather than the average. As a result of the higher thermostat setting the energy was increased.

In commercial buildings, computer suites, and hospitals, accurate control of environmental conditions is necessary, and simple thermostatic principles are seldom sufficiently precise. For such applications, modulating feedback controls are required, which need careful adjustments (78) if accurate and stable environmental conditions are to result.

The previous example cited introduces the conflicting requirements for any heating system, that is that the specified comfort conditions are achieved whilst the cost of installation and operation of the heating system is minimised. So that stable environmental conditions are obtained that satisfy the comfort conditions, the controller must be adjusted "on site" so as to match the controller response to that of the heating system.

However, the accurate adjustment of the controller is uneconomical because it is time consuming due to the slow response of heating systems, and requires specialist knowledge which the commissioning engineer does not possess. For these reasons, heating controls seldom operate at peak efficiency with respect to accurate control of comfort conditions, and as demonstrated, this may indirectly affect the energy use. In an effort to reduce the cost to the customer of installing and adjusting

the control and improve the long term performance, controls that claim to adapt to the particular heating system characteristics have been produced by several manufacturers.

Self adaptive controllers have been proposed as long ago as 1954(57), but only recently with the advent of microprocessor technology has adaptive control been considered for use in environmental control systems (11, 31, 73, 81, 86, 39, 15, 20, 25, 26, 27, 28, 29, 78). Within the sphere of environmental control, the different types of application that adaptive control has found are temperature control for a heat exchanger (11, 73), multivariable temperature and humidity control for an air conditioning unit (81), temperature control for a glasshouse (86), domestic air temperature control (29, 39, 20, 27, 28), and air temperature control for large commercial buildings (31). There have been several differing theoretical approaches to the adaptive control algorithm, however a common factor amongst all is that of parameter identification. All the controllers but one (86) use a recursive least squares algorithm to estimate the parameters of a low order discrete transfer function. The exception uses a least-squares-like gradient algorithm (87) to estimate an unknown parameter of a continuous time model of the heating system. These identification techniques can be classified as those that identify plant parameters explicitly (73, 81, 86, 31), and those which, by suitable manipulation of input/output data, identify the controller

parameters directly, the plant parameters being identified implicitly (39, 29, 27, 28, 20, 15).

The implicit schemes are based on variations of the self tuning minimum variance regulator (4), and have a great computational advantage over the explicit algorithms as a result of not having to calculate the control parameters. For the explicit algorithms, the identified plant parameters have been used to calculate the gain of a PI controller (86), derive a multivariable dead beat control law (81), and derive a state space model, from which a linear optimal control law is derived via the solution of a matrix Ricatti equation (73, 81, 31).

The implementation of complex adaptive algorithms requiring matrix Ricatti solution is possible, but requires relatively large memory capacity and advanced processing capabilities. The processing power and memory capacity of microprocessor devices are improving very rapidly, thus the cost of using time and processing intensive algorithms is becoming cheaper. However, this opportunity has been seized by many manufacturers to offer more facilities within the same product rather than making the existing product more efficient, thus for a new algorithm to be economically viable it must also be code efficient and require moderate processing power. For these reasons, considerable interest has been shown in self tuning controllers, which have been shown to be suitable for implementation on low cost microcomputer systems (28, 27, 20, 29, 39, 15).

All but one (86) of the self adaptive control schemes cited are based on a sampled data representation of the plant, controller, and if included, the disturbances. This is not surprising if one considers the popularity of the microprocessor in heating controls and the ease with which sampled data control laws may be implemented using microprocessors. A more fundamental reason for the sampled data approach is that the characterisation of noise signals is considerably more straightforward in the sampled data case, where the noise is modelled by a sequence, than if the noise is not sampled. As with all digital filters or controllers the parameters programmed into the device are inherently drift free, thus periodic readjustment of the parameters is unnecessary, and the need for time consuming calibration is removed.

A further advantage of the sampled data approach is the characterisation of time delays in terms of the Z transform. The Z transform of the time delay is an algebraic function, whereas the Laplace transform is transcendental thus making the manipulation of the system equations for control design more complex.

For most air conditioning systems the physical state of the air in the space is altered by the introduction and mixing of a volume of air of a different physical state. In a ducted warm air heating system, heated air is blown into an air space at an approximately constant volumetric flow rate, whereupon it mixes, and the resultant mixed air is exhausted at approximately the same rate as that

input. It is also likely that the input air supply has been mixed with other air supplies of differing physical states to precondition it. Thus the air mixing process is particularly important in heating, ventilating and air conditioning (HVAC) systems. However, the analytical treatment of the air mixing process is intractable for an arbitrary air space, so a small scale test rig which has some of the physical properties of a full size application was constructed, thus allowing simple models of the mixing process to be estimated and the designed controls to be tested. The test facility is described in chapter 1.2. Although this research is primarily concerned with ducted warm air heating systems, the principles derived are also equally applicable to radiator systems.

The basic functions and examples of heating systems, and the characteristics of equipment used are described in chapter 1.1. Typical heating systems are presented and the role of the temperature controller is described.

The predominant feature of self tuning control is that of parameter identification. In the majority of cases this is carried out using a recursive least squares algorithm or one of its many variants. As it is implicitly assumed that a linear model of the heating system can be estimated, then it is important to ascertain whether this assumption is justified, and if it is, then the nature of the model obtained. The modelling of a heating system similar to the test facility is presented in chapter 2.1 and a generally applicable linear model is derived. The

theoretical model derived is then approximated by a low order model obtained by practical experiment. This low order modelling procedure is described in chapter 2.2 and 2.3 and is based on cross-correlation and recursive least squares techniques. The cross-correlation results are used as a guide to the accuracy of the models identified using recursive least squares. The identification results are useful at this stage because it allows investigation of the identification technique without the complication of the control action. Important aspects that are addressed are bias in parameter estimates due to spurious data and problems caused bv non zero mean data. The results show that a linear second order model can adequatelv describe the dynamics of the test facilitv, thus making this system a suitable candidate for self tuning control. Further, that variants of the basic recursive least squares algorithm partly overcome the effects of non zero mean, noise corrupted data.

A familiar characteristic of process and heating plant models is a significantly time delayed control signal. This phenomena arises naturally in ducted warm air or radiator heating systems due to the transport delay of the heating medium, and it is introduced in modelling to simplify the model representation. In chapter 3 it is shown that the presence of time delay can seriously complicate the commissioning of a controller. Conventionally, the excess phase lag due to the time delay causes a necessary reduction in the controller gain and a subsequent

degradation in control performance, however techniques based on predictors are presented that allow controller parameters to be used as if the time delay is zero. These predictors are the Smith predictor and optimal least squares predictor.

There are several possible interpretations (38) of the self tuning controller presented by Clarke et al (15), one that is particularly useful is that of an optimal least squares output predictor combined with a conventional control law. The control law, which can be of PI form, is specified by the optimal cost function, the choice of which is not critical. The insensitivity of the system response to the choice of the control law is due to the stabilising effects of the predictor . The derivation of the self tuning control algorithm and its interpretations are described in chapter 4.1. Traditionally, self tuning controllers have used positional predictors, the inadequacy of which is shown in chapter 4.2 to be a result of non zero mean input/output data. A new form of self tuning controller (21) based on a k incremental prediction is shown in chapter 4.3 not to suffer from such inadequacy. The algorithm is summarised and necessary constants that must be specified before the controller used are described in chapter 4.4. Of particular practical importance is the performance of self tuning control on nonlinear plant and the performance when the control is given incorrect data. The effect of such practicalities is examined in chapter 4.5, results being obtained by

simulating the test facility using the low order model identified in chapter 2. The results show that the k incremental predictor form of self tuning controller has excellent offset rejection properties despite bias in predictor parameters. Also, the assignment of plant delay is shown to be a critical factor, as underestimation may cause instability.

As the practical justification for the use of self tuning control will be based on how the quality (in some sense) compares to that of the more conventional techniques, it is relevant that such a comparison be made in the early design stage. In chapter 5, the performance quality of conventional and self tuning controls is judged with respect to the ease of initial commissioning and the long term response when subjected to plant parameter variation. The comparison of the ease of commissioning is problematic as it depends on the types of commissioning procedure adopted. Thus it is more constructive to compare the sensitivity of the commissioned system response to controller and plant parameter variation. This is investigated with respect to forward gain changes. The long term performance is based on the response over 1000 samples to regular solar radiation type disturbances while the plant forward gain varies. The results were obtained from real time experiments on the test facility and show that the self tuning control is far less sensitive to forward gain changes, thus making the initial commissioning less exacting. From long term tests it was

shown that the self tuning feature of the control law allowed wide variations in plant gain to occur without a noticeable degradation in disturbance rejection. The conventional performance, however, degrades to the point of instability.

The plant discussed so far can be linearly modelled if the variation of the system variables is small, thus for small set value or disturbance changes the self tuning control law will converge to a near optimal result. There are nonlinearities within a HVAC system that cannot be linearly modelled based on the small signal assumption. Two such effects are rate and hard control signal limitations, and actuator hysterisis. These effects can seriously jeoparize a self adaptive control scheme because under certain conditions the input/output data used by the identifier will seriously be in error. The effects are investigated in chapter 6 and results obtained from simulation. The results show that for the conditions that are likely in a typical heating system then rate and hard control signal limitations will not cause a significant degradation in the response. Similarly for hysterisis, under normal circumstances, the effect of the typical level of hysterisis will not significantly affect the response. However under exceptional circumstances compensation for hysterisis is necessary, and a solution is described.

1.1 Temperature control of HVAC systems

The aim of the heating, ventilating and air conditioning (HVAC) system is to establish desired environmental conditions within the room (space) and to regulate these conditions to within specified limits. These target conditions must be achieved despite continuous and varying loads, such as heat gain from solar radiation and heat and moisture gain from occupants and the external air.

Desired environmental conditions may be specified by any, or all of the quantities of temperature, humidity, ventilation, lighting, sound level, +/- ion distribution, and other concepts more subjective and thus difficult to quantify (23). This study is primarily concerned with temperature control, the other conditions do not require such close control. For instance in the United Kingdom provided that the dry bulb temperature is between $20^{o}C$-$22^{o}C$ with air movement between 0.1 M/S - 0.2 M/S, then a variation of 40% - 70% RH will not cause a subjective feeling of discomfort (79).

There are a multitude of ways in which the temperature of the air within an enclosed space may be controlled, however it is sufficient to consider three of the most common schemes. All the schemes considered are based on the control of the sensible heat exchange between a heating coil (heat exchanger) and a supply of air of a fixed volumetric flow rate. The supply air will have been preconditioned to alter its humidity and temperature,

probably filtered, and comprise a large proportion of air previously exhausted from the room. The heat exchange process is controlled via a single temperature monitoring device, hence the process is single input, single output. This does not exclude the use of several temperature sensors, as in zone control where the temperature reading to the controller may be the average of a set of sensor readings. The heating schemes will now be presented followed by a description of control equipment that have a significant bearing on the response of heating systems.

1.1.1 Constant volume, mixed air system

The heating plant is depicted schematically by fig 1.1.1. The duct air temperature is controlled by mixing the two air streams, the position of the mixing dampers being varied as a function of the error between desired space temperature and the actual value given by 'Ts'. The hot air stream temperature 'T2' is controlled by altering the mix of supply hot water and heater return flow. This is achieved using a 3 port mixing valve which modulates the heater input between limits of 100% recirculation and 100% supply. The cold air stream is controlled in a similar way except that the cooler is supplied with chilled water, and the air leaving the cooler is controlled to have constant dewpoint. This scheme is commonly used to supply conditioned air to a number of air spaces such as a whole floor of a building.

1.1.2 Terminal reheat system

This type of heating scheme is also known as an "induction system" because the supply air induces the recirculated air across the heater, fig 1.1.2. The sensible heat input to the recirculated/induced air is altered by varying the mix of supply and recirculated water that flows through the heater. The heaters are situated within the space, for instance under windows. This type of scheme is used for individual space control and is attractive mainly because of the simplicity of installation, especially in older buildings where the previous form of heating was by radiators only. In such cases it is often possible to use the existing pipework, the supply air being drawn direct from outside through vents in the walls.

1.1.3 3 stage heating, cooling, damper system

The scheme presented is a simplified version of a 3 stage temperature control system that is used in practical applications. The plant is depicted by fig 1.1.3 illustrating the heating system and the air space. For simplicity, only the control of temperature is considered, however for such an arrangement, control of temperature and humidity is possible. The operation of the stages of heating, cooling and dampers is interlocked so that only one stage is modulating, the other two stages are at the extreme limit of their operating range. Which stage to use is determined by the control temperature 'Ts" only, thus a 'map' of temperatures can be defined which indicates

the stage to use. Such a map is shown in fig 1.1.4 for proportional control only, where the width of the stage is equal to the proportional band for that stage. The setpoint is varies such that it equals the zero output end of the stage in use. The damper setting as shown in fig 1.1.4 indicates the percentage fresh air into the conditioned space. Damper motors M1, M2, M3 are interlocked such that for maximum fresh air, damper motors M1, 3 are fully open and M2 is closed, and for minimum fresh air, damper motors M1, 3 are at the minimum setting and M2 is fully open.

The response of this type of heating system depends not only on the proportional band for each stage but also on the operating characteristics of the individual stages.

1.1.4 Final control elements: valves and dampers

The final control element (F.C.E.) is a mechanical device that alters the state of the control medium as dictated by the controller. The control medium may be water or steam in which case the FCE is some form of valve, or air in which case the FCE is some form of damper. The valve or damper characteristic is the relationship between flow through the valve or damper as a percentage of full flow, to the valve or damper travel as a percentage of that when fully open.

The important characteristic of the heating system when considering control performance is the relationship

between the heat exchanger emission and the control signal, which ideally should be a linear relationship for the full operating range. The actual relationship is affected by various factors such as the valve characteristic, design of the heating system, and the heat exchange emission characteristic. Practical heat emission characteristics are highly non-linear, thus in an effort to make the relationship between heat emission and control output linear, the valve characteristic is produced with a complementary nonlinearity. However, due to the economic need to limit the stock of valve types, manufacturers have had to produce valve characteristics that are a compromise between several conflicting requirements. Thus some overall nonlinearity is inevitable.

The design of the valve characteristic is carried out under ideal conditions, however in practice conditions are often far from ideal. In particular the valve characteristic is a function of the Valve Authority which will be different for every practical system. The valve authority is the ratio of pressure drop across the valve divided by the pressure drop across the rest of the system plus the valve pressure drop. That is,

$$A = \frac{\Delta pV}{\Delta pV+\Delta pL}$$

where the pressure drops for a two part valve are shown schematically by fig 1.1.5. The effect of varying authority on a valve characteristic that is linear for

unity authority is shown by fig 1.1.6. Thus the actual valve characteristic can vary significantly from ideal due to the influence of other parts of the heating system.

The previous examples cited in chapter 1.1 all use heat exchangers requiring precise control and in such cases three port control valves in a mixing arrangement are commonly used. The flow through the heat exchanger is approximately constant, the variation in the proportion of primary supply to the return flow, causing an alteration in the flow temperature and thus the heat emission. Fig 1.1.7 shows a conventional type of heat emission curve for constant temperature primary supply and variable primary flow. The primary supply flow to valve stroke characteristic can be altered by a suitable design of internal valve geometry or by variation in the valve authority, a range of typical characteristics is given by fig 1.1.8. The resultant heat emission to valve stroke characteristics are shown by fig 1.1.9. Curves such as (a) in fig 1.1.9 may arise as a result of incorrect sizing of heat exchanger, inappropriate choice of valve characteristic, or badly designed heating plant causing reduced valve authority. The effect on control performance of such an emission characteristic is to cause prolonged oscillation of valve position and temperature when operating at low emission values. For high emission values the controller may appear sluggish in response to external disturbances.

The damper stage in the 3 stage control scheme is used as a means of reducing the energy consumption of the cooling plant by using "free cooling". Air outside occupied buildings is often cooler or at least at a lower humidity level than within, it will also contain a higher oxygen level, thus the use of free cooling can substantially improve the subjective feeling of comfort. The mixed air system of chapter 1.1.1 uses dampers to control the flow of two streams of air of differing physical states, the relative volumetric flow rates dictating the final mixed state. Assuming perfect mixing and constant conditions then the heat energy in the mixed flow will be the sum of the heat energy in the two constituent flows. The primary concern in design of dampers for mixing systems such as described in chapters 1.1.1 and 1.1.3, is to ensure that the mixed volumetric flow rate is constant for all damper positions. The efficiency of the damper stage in 3 stage control depends on the relative environmental conditions outside and inside the building, which may not be sufficiently different to make accurate control possible. Thus in practice the damper stage is usually controlled with a small proportional band controller, such that the dampers are commonly either fully open or closed. The mixed air system of chapter 1.1.1 commonly supplies air to a zone. The application of certain parts of the zone may require that the air is further conditioned, such as in the case of a computer suite or a hospital operating theatre where accurate environmental conditions are required. The accurate control of the duct conditions 'T2', 'T1' (see fig 1.1.1) can be achieved by using

integral control. This type of control is only practical when the time lag of the plant is very short, which is the case for this particular system.

The valves and dampers require some form of actuation to alter the valve stem or damper blade position as dictated by the controller. These can be based on an electric motor or a pneumatic thruster, whichever power source is the most convenient. For all practical purposes, actuators have a linear steady state relationship between control signal and valve stem or damper blade position. The time taken for the actuator to alter its output by 100% is known as the stroke time, which may be in the range 20-360 seconds. Both electric and pneumatic actuators exhibit hysterisis. This is the time delay between a change in direction of the control signal and the corresponding change in direction of the valve stem or damper position. Typical values of hysterisis are likely to be less than 2.0% of stroke time nominally, however this may increase due to wear.

1.2 Description of experimental equipment

So that a temperature controller can be designed, it is necessary to obtain a simplified linear model of the thermodynamics of the heating system and building. In most cases this requires that the thermodynamics of the air in the space be modelled, however due to the intractable nature of the air mixing process, this has

necessitated gross simplification of the governing equations to be made. The accuracy of such manipulations is questionable, so a test facility that exhibits the same characteristics as a practical air mixing system, is useful as a means to test the accuracy of the simplified models, and to prove the effectiveness of controller design based on these models.

The test facility comprises an enclosed air space which is supplied with heated air at an approximately constant volumetric flow rate. Temperature monitoring and control of heat input to the air is carried out by a general purpose microcomputer and the necessary interface equipment. The enclosed air space is situated within an intermittently heated laboratory, thus the internal air temperature will be affected by the laboratory air temperature as a result of heat conducted through the walls, and also from the supply air which is obtained from the laboratory itself. This situation then closely resembles a practical application where the internal temperature of a building is disturbed by the variation in external air temperature. A recording of the laboratory air temperature was taken over a one week period and is shown by fig 1.2.1. Other disturbances on a practical heating system, which are due to direct solar radiation and rapid changes in occupancy, can be simulated in the test facility by causing step changes in the heat input to the air space. This is facilitated by switching on a 60 watt incandescent lamp within the air space.

The sections to follow give a more detailed description of the individual parts of the test facility.

1.2.1 Microcomputer

The microcomputer used is a general purpose device based on the 8 bit Z80 microprocessor, with 32768 bytes of random access memory (RAM) which is used for data storage and running programs, and 32768 bytes of non-volatile read only memory, part of which is used to load the CP/M operating system into RAM after a reset. The operating system plus data and programs are stored on a double sided, single density 8" floppy disk unit.

The control and analysis programs were written in Z80 assemebler and assembled using the MACRO-80 macro assembler program. Although the writing of assembly language routines is initially time consuming, once a suite of the most frequently used subroutines are written, then further program development can proceed quite rapidly. For this study floating point number representation will be used. Arithmetic subroutines for the floating point number representation are widespread especially for the 80 80 microprocessor. The routines presented by Cope (22) will be used as they were developed especially for self tuning control studies (15) and comprise all the necessary arithmetic operations. These consist of addition, negation, multiplication, division and square root estimation.

The floating point operands are stored in memory as 3 consecutive bytes in the order, exponent, least significant byte of mantissa and most significant byte of mantissa. The number range is from $\pm 0.5 \times 2^{-64}$ to $\pm(1-2^{-15})x2^{63}$, which is $\pm 2.7105x10^{-19}$ to $\pm 0.9223x10^{19}$, and accuracy is one part in 2^{15}. Overflow results are set to $\pm(1-2^{-15})x2^{63}$ and underflow results to zero which is represented by all 3 bytes equal to zero. This then ensures that zero is less significant that the smallest number represented in the number system. The accuracy of the arithmetic operations is equal to that allowed by the number representation.

The microprocessor communicates with the interface circuitry via 2, 8 bit, addressable memory locations which reside in an on-board PIO (parallel input/output) chip.

One of the ports is configured as outputs, designated AO-A7, bits AO-A6 are for heater control and A7 for disturbance lamp control. This then gives a potential range of control signals from 0 to 127 and a resolution of 1 part in 128. The other POI port, designated BO-B7, is configured as 4 inputs B4-B7, and 4 outputs BO-B3. 4 inputs are used for the ADC (analogue/digital converter) as the temperature reading is 4, 4 bit BCD digits. The outputs are used to control the multiplexing of ADC digits and the analogue multiplexer.

1.2.2 Air space

The air space is constructed as a cube of side dimension 1.3m, giving an approximate volume of $2.2m^3$. The walls of the cube are constructed from 25mm polystyrene bonded to hardboard sheet. To allow access to the space, the floor is covered with a platform,and placed at various positions in the space are temperature sensors which are connected to the interface board, the exact layout is shown by fig 1.2.2.

So that heated air evenly mixes with the air in the space, the air is introduced at the top of the cube and exhausted at the base. The supply duct houses the fan, heater, and damper, and is of cross-sectional area $0.0103m^2$. Theoretically the fan should supply $46m^3/hr$, however due to exfiltration and back pressure, the measured flow rate at the exhaust is reduced to $34.4m^3/hr$. This is equivalent to 15.6 ACH (air changes per hour), which is relatively high compared to that for an office building of typically 5 ACH. However, scaling down the supply rate to such an extent causes problems with fan selection and disturbance effects of natural ventilation within the laboratory. The essential factor is that the test facility exhibits the characteristics of an air mixing process which are similar to that found in a practical situation.

1.2.3 Heater and control signal interface

Considering the size of the air space the most economical way of heating the air is by electrical means. This is achieved by controlling the power dissipated in the supply duct by a lattice of resistors. These resistors cause insignificant restriction to the passage of air through the duct. So that the temperature of the air space can be controlled at a fixed value despite the variation in supply air temperature (see fig 1.2.1), then a maximum possible control range of 8^{o}C above ambient was chosen as being sufficient to cope with most climatic conditions. The most convenient way to supply the power to the heater resistors to cause such an increase in temperature is to connect them in parallel across 240 Vac, in which case 20, 6.8kΩ resistors are required, giving a maximum power dissipation of 169 Watts.

To control the average power dissipated, the voltage is applied to the resistors for a controllable percentage of each cycle of the mains voltage. This percentage corresponds to the control signal output from the microcomputer . As the period of one cycle is very short compared to the time constant of the temperature response of the heater resistors, then the temperature of the resistors is effectively constant for a constant control signal.

The mains voltage is switched across the resistors by a triac during every half cycle of the mains voltage. The triac can be made to conduct at any of approximately 128 equispaced points between the start and end of a half cycle. This is achieved by generating a synchronised waveform of 256 times the mains frequency using a phase locked loop (PLL) and zero crossing detector (ZCD) arrangement. The PLL output frequency is used to clock a shift register producing an increasing binary count, which is reset by the ZCD every half cycle. This binary output from the shift register is then compared with the binary output from the microcomputer by means of a digital comparator, and when the comparator signals equality then the triac is energised. Thus the binary control signal actually determines the percentage of time for which the triac is switched off, so before the control signal is output it must be subtracted from 127. The detailed circuit diagrams of the heater control and disturbance lamp control circuits are given by figs 1.2.3a, b, c, d, e.

The resolution of the control signal is one part in 128, however the practical range is limited to 10 to 120 due to the pulse width of the ZCD signal. Each half cycle of the mains voltage will be subdivided into 128 parts, each increment being denoted as 1^{o} firing angle of the triac. The control signal is then in units of degrees firing angle.

1.2.4 Temperature monitoring interface

The temperatures sensors used are room temperature detectors supplied by Satchwell Control Systems Ltd, type DRT 2453.

The detector consists of a negative temperature coefficient thermistor, a network of resistors to help linearise the thermistor characteristic and a housing that allows wall mounting. The detector is designed for use over a range of -5°C to 40°C which covers the temperatures of interest.

To convert the thermistor resistance variation to a voltage that may be input to an ADC, and to reduce the self heating effects on the thermistor, it is common to encorporate the detector in one arm of a Wheatstone bridge circuit, as shown by fig 1.2.4. The potentiometer 'R_Z' is set to give zero 'Vin' when 'R_D' is equal to the resistance corresponding to 10°C, which is the minimum of the chosen temperature range, the maximum is 40°C. The variation in 'Vin' is amplified by an inverting amplifier to give 0V-2V output for a change in detector temperature of 10° - 40°C. This voltage then forms the input to the ADC, the digital output of which has a numerical range of 0-2000 for an input voltage range of 0V-2V. Furthermore, so that several detectors may be read without unnecessary duplication of hardware, an analogue multiplexer is used to connect the amplifier to several detector bridge circuits. The detailed circuit diagram of the temperature

monitoring circuit is given by figs 1.2.5a, b.

As there are two PIO output bits to control analogue multiplexing then four detectors can be read, which is adequate for this study. The conversion time of the ADC is 250mS, thus four detectors can be read in approximately 1 second. To allow for tolerances on components, especially the input filter capacitor 'C', then the analogue multiplexor is switched every 500mS.

This then gives rise to an approximate two second delay to read the detectors, which is insignificant for the majority of HVAC control applications, and this study in particular.

As the temperature/resistance curve of the detector is nonlinear then the curve relating temperature to ADC output is also nonlinear. The temperature/ADC output curve was obtained by placing a fixed resistor in place of the detector, the value of which corresponded to the detector resistance at the given temperature, and the curve is given by fig 1.2.6. So that the temperature can be read to a reasonable degree of accuracy, and the complexity of the circuitry is minimised, it is common to piecewise linearise the curve over fixed temperature intervals. A $1^{o}C$ error between measured temperatures and actual temperature is likely to be sufficiently accurate for this study, thus a single linear curve, as shown by fig 1.2.6, can be fitted to the actual curve to allow

measurement of temperature from ADC output.

Further errors arise as a result of spurious variations in measured temperatures, these effects being known collectively as measurement noise. These variations may be due to electrical signals coupled to the detector leads or signals generated within the controller itself. Also, the actual temperature measured by the thermistor may not be representative of the temperature of the air surrounding the occupants, this being due to the inappropriate positioning of the detector or heat input disturbances direct to the detector.

The electrically generated measurement noise can be measured by reading the temperature with a fixed resistor in place of the detector, the result of such an experiment being given by fig 1.2.7. The variation of the measurements is equal to $\pm 0.015^{o}C$ about a fixed value. This variation is equivalent to ±1 LSB (least significant bit) of the ADC and thus this is the minimum noise variation possible without further filtering of the measured value.

To avoid an error occurring between measured temperature and the temperature most representative of that experienced by the occupants, then the positioning of the detector within the space must be done with care. For instance, the detector must not be placed in direct sunlight, or immediately above a radiator, or below a duct

outlet, otherwise a false reading will be given. This phenomenon can be demonstrated by measuring the temperature response of the air space at two different positions within the test facility. The variation in air temperature is in response to a step change in heat input, and the detector positions are at 'A' and 'B' as shown by fig 1.2.2. Because the time response of the test facility is fast relative to a practical application, it was considered prudent to remove the thermistors from the detectors on 'flying leads', so that the thermal mass of the housing would not affect the measured response. The time constant of the thermistor is 10 seconds maximum in still air and 1 second maximum in stirred oil (90). The response for the detector in position 'A' and 'B' are given by fig. 1.2.8.

The difference in time constants of the two step responses is approximately 400 seconds, which cannot be due solely to the air velocity as the time constant in still air is only 10 seconds maximum, thus the difference must be due to the pattern of air mixing within the space. In practical applications it is common to place the detector in the exhaust duct because this gives a reading that is closest to the average temperature within the space. This average value is close to the actual temperature at any point if the air is efficiently diffused throughout the air space.

2 Modelling and identification of the heating system

The modelling of HVAC systems is carried out to different levels of complexity depending on the intended use of the model. Models range from the very complex, for use in prediction of energy demand and design of plant (56), to the most simple dynamic form useful for the design of control loops (48).

The aim of the modelling procedure to be presented is to obtain a sufficiently simplified mathematical description of the heating system so that a controller may be designed to satisfy both performance and economic constraints. This aim may be achieved by the detailed analysis of the system to obtain the governing equations, and perhaps the linearisation of these equations about system operating points. The equations derived however will apply to only one particular plant, and due to the complex derivation will be difficult to apply generally. Alternatively, by gross simplification of the system equations a simple model can be derived, however the accuracy of such a model would be doubtful.

A powerful alternative approach for modelling systems is to analyse the system equations so as to obtain the structure of a simplified model, and then from experimental data obtain estimates of the unknown coefficients of the structures in such a way as to minimise a modelling error criterion. This

type of approach is used in the following sections to develop a model of the heating system.

The derivation of the significant dynamic structure of the heating system is described in chapter 2.1. This is achieved in systematic manner by simplifying the derived governing equations of each of the individual stages of the entire system. The resultant structure is shown to be a special case of a more generally applicable structure which can be simplified to a form convenient for controller design.

The techniques used to estimate the unknown coefficients of the simplified structure are described in chapter 2.2. These techniques comprise cross-correlation and generalised recursive least squares estimation using a psuedo-random-binary-sequence (PRBS) input.

The results of cross-correlation and generalised least squares identification is compared in chapter 2.3.

The techniques of cross-correlation and recursive least squares estimation are developed as a means of obtaining the 'best' coefficients of a given model in a proper optimal sense. The self tuning controller presented in chapter 4 is based on a parameter estimator utilising the recursive least squares algorithm described in this chapter. The initial modelling studies were helpful in developing the parameter estimator as several difficulties

were highlighted associated with the implementation of the algorithm.

This study does not attempt to derive a model for the human conception of comfort conditions. Such a model could be used as a way of judging the degree of human satisfaction in the controlled environmental conditions. Although this approach has been suggested (85) as a necessary element in the modelling procedure, such subjective models are not easily incorporated into the design procedure. The criterion of mean square error from set point can be used as a means of judging quality of temperature control because it can readily be incorporated in the control design procedure, and also relates well to other more subjective criteria derived from experiment (34).

2.1 Derivation of the heating system model

The modelling technique to be presented is to obtain the most general linear transfer functions possible for all the system components and thence to obtain a generally applicable structure. To do so, few assumptions are made as to the significance of the system components on the overall response, this being necessary because of the variable nature of the application. The general system structure derived is shown to be a special case of a structure which can be modelled in an optimal sense by a reduced order transfer function that is amenable to

control design. Having defined the reduced order model it is then necessary to carry out experiments so that the coefficients of the model can be estimated. This is the subject of chapter 2.2.

2.1.1 The heat exchanger and hot water valve

The heat exchanger is fed with hot water via the mixing valve 'V1' (fig 2.1.1) which mixes the boiler output with a portion of return flow from the heat exchanger. The modelling procedure will consider the combination of valve and heat exchanger as a single stage, such that the input is the small perturbation of valve position from mean, 'ΔV1', and the output is the small perturbation of heat exchanger output air temperature from mean, 'Δθ1'.

The heat exchanger is known to be a nonlinear, multivariable, distributed parameter system (41), and without significant simplification control design is impractical. However the results of experimental studies (70) show that the significant dynamics can be approximated by a Laplace transfer function of the following form.

$$\frac{\Delta\theta 1(S)}{\Delta\theta 2(S)} = G1(S) = \frac{K1.\exp(-S.D4)}{(1+T1.S)(1+T2.S)} \qquad (2.1.1)$$

where 'Δθ2' is heat exchanger supply temperature change. This relation was obtained for heat exchangers commonly used in this application and is based on the assumption of constant heat transfer coefficients. As

heat transfer coefficients are dependent on fluid flow rates (56), then it is essential that the flow rates are constant and this requires that the valve is correctly matched to the system (23). Assuming that the valve is correctly chosen then for small variations in valve position it is possible to relate heat exchanger input, '$\Delta\theta 2$', and valve input, '$\Delta\theta 1$' by a constant gain. In which case the valve and heat exchanger dynamics can be modelled by equation 2.1.1.

The approximate dynamic form described is commonly used for heat exchanger control design and will be shown by experiment to be a good approximation of the dynamics of the test facility heat exchanger.

2.1.2 Valve actuator

The valve actuators to be considered are of the modulating type, in that the valve position is uniformly variable between fully open and fully closed. Both electrical and pneumatic power sources are commonly used, the distinction between types of actuator is in how it interfaces to the controller. The two forms of modulating actuator commonly used are the self-positioning and incremental types. Self positioning actuators position the valve stem proportional to the control signal magnitude, this being achieved by utilising an error amplifier and feedback potentiometer to indicate

valve position. The incremental actuator causes a change in valve position that is proportional to the control signal unless the valve reaches its limits. An all electrical incremental actuator comprises of a constant speed motor and reduction gearbox which can be signalled to increase or reduce valve position at a constant rate. Because the actuator rate is constant the change in valve position is proportional to the length of time the control signal is present.

The use of incremental control is becoming more popular with digital control (7) because the control output is merely a tristate signal, that is, the control signal is an indication for the actuator to increase, reduce, or not to change the valve position. The disadvantages of incremental control are that the controller has no direct indication of valve stem position, and the control signal is limited by the sample time. This can cause degradation of control performance and will be discussed further in chapter 6.

The action of the incremental actuator is to accumulate control signals so that it appears in the control loop as a discrete integrator, and can be approximately modelled by the Z domain transfer function.

$$A(Z) = \frac{Ka.Z}{(Z-1)} \qquad (2.1.2)$$

The gain term 'Ka' is proportional to the slew rate which is given by the reciprocal of the time taken for the value to alter its position by 100%, typical values lie in the

range, 1/60 S^{-1} to 1/360 S^{-1}.

2.1.3 Ductwork

The ductwork comprises all the ducts and mixing chambers after the heat exchanger and before the air space. Depending on the exact nature of the control scheme the heat exchanger may be quite remote from the air space, in which case if it is assumed that the air supply rate is constant then the ductwork will introduce a fixed time delay in the control loop. It is common practice (92) to account for this delay plus the dynamic effects of a number of small time constant exponential lags by a pure time delay. Any loss or gain of heat and/or air by conduction, infiltration, or exfiltration will be assumed negligible while within the duct, hence the ductwork can be modelled by the laplace transfer function.

$$D(S) = \exp(-S.D2) \qquad (2.1.3)$$

where 'D2' is the time delay in seconds.

2.1.4 Air space and building fabric

The air is supplied via the duct to the controlled environment where after mixing with the space air is exhausted through the exhaust duct. The exact nature of the mixing of the air is extremely complex and it is widely accepted that for control design, instantaneous-perfect mixing must be assumed. In which case the temperature and humidity are the same at all points in

the space at any one time. Given this initial assumption then the dynamics of temperature and humidity within the air space may be derived by considering the material balances of dry air and water together with an enthalpy balance of the moist air. The results of such an analysis (66) show that by linearising the governing equations by considering small perturbations of the variables about an operating point, then the temperature and humidity are given by uncoupled first order differential equations. The equations can be expressed in matrix form as follows.

$$\dot{\underline{X}} = A.\underline{X} + B.\underline{U} + D.\underline{V} \qquad (2.1.4)$$

where

$$\underline{X} = \begin{Bmatrix} \Delta h1 \\ \Delta\theta 3 \end{Bmatrix}, \quad \underline{U} = \begin{Bmatrix} \Delta h2 \\ \Delta\theta 4 \end{Bmatrix}, \quad \underline{V} = \text{disturbance input vector}$$

$$A = \begin{Bmatrix} a11 & 0 \\ 0 & a22 \end{Bmatrix}, \quad B = \begin{Bmatrix} b11 & 0 \\ b21 & b22 \end{Bmatrix}, \quad D = \text{disturbance parameter matrix}$$

$\Delta\theta 4$ = change in duct outlet temperature

$\Delta\theta 3$ = change in air space temperature

$\Delta h1$ = change in air space humidity

$\Delta h2$ = change in duct outlet humidity.

The thermodynamic system has a convenient electrical analogy (75), fig 2.1.2, for which the following analogies apply, (for constant humidity).

ELECTRICAL QUANTITY		THERMAL QUANTITY	
I CURRENT	(amps)	h HEAT FLOW RATE	(kg-cal/sec.)
V VOLTAGE	(volts)	θ TEMPERATURE	(^{o}C)
R RESISTANCE	(ohms)	R RESISTANCE	(^{o}C.sec/kg-cal)
C CAPACITANCE	(farads)	C CAPACITANCE	(kg-cal/^{o}C)

The equations (2.1.4) were derived assuming that the building envelope is light and well insulated such that the effects of thermal capacitance due to the building and contents is insignificant. In general the effect of this thermal capacitance will be significant, indeed it may be the predominant factor (45). The equations describing heat flow through an homogeneous material are partial differential but may be approximated by a set of total differential equations. This approximation is a consequence of assuming the material is a multilayer slab and heat flows are perpendicular to the slab surface. The electrical analogy can once again be invoked to help describe the heat flow through the material, fig 2.1.3. This type of representation is commonly used (47) to find the temperature response of an homogeneous material. If each resistor/capacitor stage is considered a two port network then the total network transfer function can be found by a cascade of the ABCD parameter matrices. The result of such an analysis reveals that the transfer function relating air space temperature perturbation to duct outlet temperature perturbation is given by an equation of the following form,

$$\frac{\Delta\theta 3(S)}{\Delta\theta 4(S)} = \frac{K2.(1 + \sum_{i=1}^{N-1} S^i.b_i)}{1 + \sum_{j=1}^{N} S^j.q_j} \qquad (2.1.5)$$

where 'N' is the number of slab layers plus one. This is a very general description of the dynamics of temperature within an enclosed air space with input equal to supply air temperature and no assumptions are made at this stage as to the significance of the roots of 2.1.5.

2.1.5 Temperature Transducer

The commonly used form of temperature transducer for environmental control is the negative-temperature-coefficient thermistor, this is due to its low cost and robust nature. The thermistor has a nonlinear resistance-temperature characteristic which can be linearised to some extent by a resistor compensating network. Use of microprocessor-based controllers (74) has made it possible to further linearise the detector characteristic by use of a 'linearising' table stored in the controller's non-volatile memory.

Positioning the thermistor has been discussed in chapter 1, and so as not to alter the air space measured response the thermistor will be placed in the exhaust air stream as described.

2.1.6 Modelling the disturbances

Disturbances to the heating system are from both internal and external sources. The external source comprises solar radiation, wind, and ambient effects. Solar radiant heat incident at the outside surfaces of the building is partly transmitted to the interior by conduction through opaque materials. This absorbed radiation has the same effect as an increase in the outside temperature. The calculation of this heat gain is facilitated by the concept of 'sol-air' temperature, which is defined as the outside air temperature which in the absence of solar radiation would give rise to the same temperature distribution and rate of heat transfer through the opaque structures as exists with the actual outside air temperature and solar radiation.

The internal source comprises direct solar heat gain through windows, heat gain from occupants, machinery etc, and electrically coupled measurement noise.

The exact nature of the disturbances is a function of the siting of the building, the building application, and the area of the windows, and hence the effects of such disturbances are difficult to quantify. Hence in an effort to generalise the disturbance effects the disturbances are assumed to be generated by a deterministic and a stochastic source. The deterministic disturbance is to account for sol-air effects and can be modelled by a

step, ramp or sinusoidal function. The stochastic disturbance accounts for most of the high frequency effects and is modelled by autocorrelated white noise. This noise process is useful because it is sufficiently general to represent any random process (2), and it allows controls to be readily designed to compensate for such disturbances. This disturbance model structure is similar to that used in a recent study (27) of adaptive space heating systems.

2.1.7 Approximate dynamic modelling of heating system

The transfer function relating valve position perturbation to air space temperature perturbation can be obtained by combining the transfer functions of the individual elements, such as the heat exchanger, duct, etc. Hence combining equations 2.1.1, 2.1.3, 2.1.5 the following transfer function is obtained.

$$\frac{\Delta\theta 3(S)}{\Delta V1(S)} = \frac{K1.\exp(-S.D1)}{(1+T1.S)(1+T2.S)} . \exp(-S.D2).K2.\frac{(1+\sum_{i=1}^{N-1} S^i.b_i)}{(1+\sum_{j=1}^{N} S^j.a_j} \qquad (2.1.6)$$

This equation is a special case of the general transfer function given by

$$\frac{\Delta\theta 3(S)}{\Delta V1(S)} = K.\exp.(-S.D).\frac{(1+\sum_{i=1}^{M-1} S^i.c_i)}{(1+\sum_{j=1}^{M} S^j.d_j)} \qquad (2.1.7)$$

In the same way as the complex dynamics of the heat exchanger were approximated (70), this equation can be approximated by a reduced order model (40,10) of the same form as equation 2.1.1. The method of moments described by Gibilaro and Lees (40) for determining the best (in some sense) values for the parameters of equation 2.1.1 is most suitable for matching the responses of theoretical models, this is due to the practical difficulty of measuring moments of experimental responses. In the next chapters, techniques will be described for accurately determining the best values, in the least squares sense, of the reduced order discrete model given by

$$G(Z^{-1}) = \frac{Z^{-k}.(b_1.Z^{-1}+b_2.Z^{-2})}{(1+a_1.Z^{-1}+a_2.Z^{-2})} \qquad (2.1.8)$$

from experimental data. This equation is the discrete equivalent of the continuous second order delay transfer function given by equation 2.1.1.

The use of such a model for control design has been found to be adequate for numerous applications and will be used throughout the control design.

2.2 Identification of linear systems

The two identification techniques to be presented are, weighting sequence identification by cross-correlation and Z transform transfer function identification utilising generalised least squares theory. The weighting sequence and Z transform transfer function being equivalent

representations of a discrete system.

Identification of the weighting sequence by crosscorrelation has the advantage that it requires little a priori knowledge of the plant and noise dynamics and is easily obtained in the presence of correlated disturbances. However control design is more easily handled with the transfer function representation. As obtaining the transfer function from the weighting sequence is rather cumbersome, then more direct methods are usually employed. Such a direct method is that of generalised least squares (GLS) which utilises a least squares (LS) estimation algorithm. In later control studies, the LS algorithm is used as a parameter estimator and so its initial development in this section is valuable to establish a familiarity with the technique and to debug the computer programs to be used later.

The input signal to the plant for identification purposes is a constant mean value plus a PRBS perturbation. This input allows estimation of the weighting sequence by crosscorrelation without deconvolution (52) and results in estimates that are optimal in the sense of least error variance (13), furthermore a PRBS is a useful test signal for least squares identification (52). Hence, although the weighting sequence does not lead to straight forward control design it is an attractive means of comparison of accuracy for the transfer function model. The accuracy of the transfer function estimate is judged

in this case by how closely the weighting sequence resembles the unit pulse response of the transfer function.

The basic theory of weighting sequence identification by crosscorrelation is described in chapter 2.2.1 and the planning of PRBS input experiments is outlined in chapter 2.2.2. Planning PRBS tests requires knowledge of the process which is conveniently obtained by a step response test.

Chapter 2.2.3 describes the practical crosscorrelation experiments conducted on the test facility in which major problems due to ambient temperature drift and plant nonlinearity are highlighted and solutions derived.

The identification of parameter models using least squares theory is described in chapter 2.2.4. The algorithms presented are suitable for implementation using short word length number representations such as would be appropriate for a microprocessor controller. The two major problems in implementation of the identifier are shown to be nonzero mean data and autocorrelated disturbances. The former problem is avoided by filtering the data, however this correlates the noise. To overcome bias in estimates due to autocorrelated noise generalised-least-squares (GLS) theory is described in chapter 2.2.5, and two GLS algorithms are presented.

2.2.1 Weighting sequence identification using crosscorrelation

The crosscorrelation technique is very well known and reported in many studies (60, 91, 42), hence only the main results will be stated here.

The weighting sequence is the sampled output of a previously relaxed system excited by a unit pulse input. The output can be given in terms of the weighting sequence, for any discrete input, by the convolution summation as follows:-

$$y(t) = \sum_{i=1}^{N} W_i.u(t-i) + d(t) + e(t) \qquad (2.2.1)$$

where 'W_i' is the weighting sequence, 'u(t)' is the sampled input, 'd(t)' is a constant or very slowly time varying disturbance, and 'e(t)' is a member of an autocorrelated noise sequence.

If the sample time is 'ST' then to avoid overlapping of data,'N' should be given by the following relation:-

$$ST.N > \text{settling time of the system} \qquad (2.2.2)$$

The Z transform of the weighting sequence is given by:-

$$W(Z^{-1}) = W_1.Z^{-1} + W_2.Z^{-2} + ... W_N.Z^{-N} \qquad (2.2.3)$$

This is equivalent to the Z transform transfer function of the stable system, such that:-

$$W(Z^{-1}) \cong \frac{B(Z^{-1})}{A(Z^{-1})}$$

Therefore, given the transfer function the weighting sequence can be obtained by division of '$B(Z^{-1})$' by '$A(Z^{-1})$'.

For an output given by a PRBS of magnitude 'a' the autocorrelation of the input is given by the following:-

$$\phi_{uu}(k.ST) = \frac{1}{N}\sum_{i=1}^{N} u(i).u(i+k) = a^2 \text{ if } k = 0$$
$$= -\frac{a^2}{N} \text{ if } k \neq 0 \qquad (2.2.5)$$

where 'N' is the length of the PRBS, which is also taken as the number of elements in the weighting sequence. For such an input the weighting sequence is given by the crosscorrelation of input/output data for an integer number of sequences. The expectation of the crosscorrelation is given by the following relation (91),

$$E\{\phi_{yu}(k.ST)\} = W_k.a^2.\frac{(N+1)}{N} - \frac{a^2}{N}.\sum_{i=1}^{N} W_i \pm d.a \qquad (2.2.6)$$

The weighting sequence estimates are then proportional to the expectation minus any constant offset. The error variance of the estimates from the actual values is given by

$$\sigma_w^2 = \frac{\sigma^2}{p.a^2.(N+1)} \qquad (2.2.7)$$

for white noise disturbance

where 'p' is the number of sequences and 'σ^2' is the noise variance. The variance is inversely proportional to the number of sequences, hence for low signal to noise power

ratio, a large number of sequences may be required. This value of variance (2.2.7) applies for both uncorrelated and correlated noise sequences (13) and is optimal in the sense of minimum variance. Hence crosscorrelation with a PRBS input gives an estimate of the weighting sequence, the accuracy of which is the best that can be obtained from the available data. The perturbation magnitude can be made very small if the number of sequences is large or the noise magnitude is small, and the ambient conditions do not significantly change. In which case the linear model is more valid than that obtained by tests using large inputs such as sinusoids or steps.

2.2.2 Planning crosscorrelation experiments

In practice a successful crosscorrelation experiment will require the application of several pseudo-binary-sequences, in which case the experiment time will be relatively high. Therefore, rather than set up the experimental conditions by trial and error, it would be advantageous if the conditions could be estimated before experimentation. The parameters that must be fixed are as follows:

(i) the sequence length 'N'
(ii) the sample time 'ST'
(iii) the perturbation magnitude 'a'
(iv) the number of sequences to apply 'p'

If the PRBS is applied during normal operation of the plant then the fluctuation of the output may place a limit on

magnitude of the perturbation 'a'. This is not a restriction with the test facility so the only limit on 'a' is that it is small enough to preserve linearity. The range of 'a' for which the plant can be considered linear can be found by recording the output steady state value for various values of input throughout its range. The predominant nonlinearity in the plant is the heat exchanger, so to ensure linearity it is sufficient to establish that the duct temperature varies as a linear function of control signal. This indicates the steady state linear region, the effect of dynamic nonlinearity will be dealt with in chapter 2.2.3. The steady state difference in duct and ambient temperature is plotted against control signal in fig 2.2.1. From thus plot it is evident that the duct temperature varies linearly for control signals in the range of 35% to 65% of maximum.

The time for one sequence must be greater than the settling time of the system to avoid overlapping of data, and this can be conveniently estimated from the step response. The response of exhaust temperature to a change in control of 100% is shown by fig 2.2.2. The response is dominated by an exponential of time constant 400 seconds, hence the settling time will be a minimum of five time constants, that is, 2000 seconds. The variation of the weighting sequence estimates in the settled region is an interesting feature as it serves to indicate the presence of asymmetrical nonlinearities, hence initially a sequence length of approximately 2500 seconds will be used. To obtain sufficient resolution a sequence length

of 63 cycles is usually sufficient, in which case the sample time is approximately 40 seconds.

The number of sequences to apply is a function of the signal/noise ratio and also the inevitable plant drift. This is because the low signal/noise power ratio requires that a large number of sequences are applied, in which case the resultant long experiment time may cause drift to become significant. As drift is difficult to predict beforehand, as are all external disturbances, an arbitrarily large,20,number of sequences were applied and the resultant input/output data stored on floppy disc. This allowed the crosscorrelation to be carried out off-line for as many sequences as necessary, and more significantly any of the 20 sequences could be chosen, hence allowing the part of the experiment where drift was minimal to be used.

It is interesting to consider the step response of fig 2.2.2 compared to that predicted by models used in previous HVAC control studies. The most popular form of model for control design (75) is the time delay plus time lag model given by the following transfer function.

$$G(S) = \frac{K.\exp(-S.D)}{(1+S.T)} \qquad (2.2.8)$$

The study due to Harrison et al (48) is representative of many in which the time constant 'T' is estimated to be equal to the fill time of the space. The fill time of the cube can be estimated by measuring the exhaust air flow

rate and the volume of the space and is given in this case by,

$$T = \frac{V}{q} \cong 230 \text{ seconds} \qquad (2.2.9)$$

where 'V' is the volume and 'q' the volumetric flow rate. This estimate does not compare well with that obtained from the step response, fig 2.2.2, the error being attributed to the presence of other thermal mass besides air in the air space, and imperfect mixing.

The time delay 'D' accounts for transport lags and many small time constant lags (92) and is consequently difficult to predict, however Harrison et al (48) suggests that an estimate for 'D' is the fill time of the space. Comparing the effective time delay 'Td' to the fill time shows that this is a poor estimation procedure, for this particular plant.

The steady state gain 'K' is difficult to estimate in practice because it is operating point dependent, and also a function of the configuration of the plant, which is outside the influence of the control engineer. Hence even the simplest plant has a transfer function that is difficult to predict, and some form of experimentation is required to obtain a reasonably accurate model.

2.2.3 Crosscorrelation experiments

The input-output data for crosscorrelation requires to be zero mean, thus before mean values can be calculated and

crosscorrelation started, the plant must be allowed to reach steady state, so that perturbations can be calculated.

Once steady state has been reached then the perturbations can be calculated assuming that the mean values remain constant or some compensation is included for any variation. The main cause of mean value drift is the variation in ambient temperature and as this is measurable then compensation can be introduced. The frequency of the variation of the ambient temperature can be assumed to be much lower than the break frequencies associated with the test rig, such that any variation in ambient will only significantly affect the mean values of duct, exhaust and space temperature. Therefore drift can be compensated for by merely adding any change in ambient temperature to the mean values of the other temperatures.

To demonstrate the crosscorrelation technique and the ambient temperature compensation scheme, a crosscorrelation experiment will be carried out on the duct temperature. The duct temperature was used as the system output in the initial studies because the response is at least ten times faster than for exhaust hence initial debugging of the programs and theory required the minimum of time. For the experiment, the input control signal is given by,

$$u(t) = 50\% \pm 8\% \text{ of maximum}$$

and the measured output is given by,

$$y(t) = TD(t) - TD(0) \qquad (2.2.10)$$

$$y(t) = TD(t) - \{TD(0) + (TA(t) - TA(0))\} \qquad (2.2.11)$$

where 'TD(t), TA(t)' are the duct and ambient temperature at time t, t = 0 refers to the initial temperatures at the start of the experiment which are the assumed initial mean values. The crosscorrelation was carried out for the two types of output given by equations 2.2.10, 2.2.11 for 9, 63 stage PRB's, with sample time of 12 seconds. The effect of the drift compensation scheme can then be judged by comparing the weighting sequence estimates for the two outputs. From equations 2.2.6 and 2.2.5 the estimated weighting sequence is given by

$$\text{Est } \{W_i\} = \frac{N}{(N+1).a^2} \cdot \{\phi_{yu}(k.ST) - \text{mean of settled crosscorrelation curve}\}$$

and the estimates are shown by fig 2.2.3a and fig 2.2.3b.

Although the primary peaks in the curve are almost identical, the variation of the estimates in the settled portion is drastically different. The secondary peaks for the uncompensated output are uncharacteristic of thermal systems and hence may be attributed to drifting ambient conditions.

With this drift compensation and the guide lines laid down in section 2.2.2, crosscorrelation can be carried out between exhaust temperature 'TX' and control input. The results of crosscorrelation for input given by,

$$u(t) = 50\% \pm 12\% \text{ of maximum}$$

and output given by,

$$y(t) = TX(t) - (TX(0) + TA(t) - TA(0))$$

is given by fig 2.2.4. The variation of the estimates in the settled portion is dominated by secondary peaks at 36, 55 and 60 samples which are not reduced for an increase in input perturbation magnitude of 19% as shown by fig 2.2.5. In previous crosscorrelation studies (60, 43) the presence of secondary peaks has been attributed to asymmetrical nonlinearities in the plant due to different heating and cooling rates. Although these nonlinearities are unlikely to cause difficulties with control it is interesting from the modelling point of view to account for the effects.

A technique for reducing the magnitude of secondary peaks in crosscorrelation functions due to asymmetrical nonlinearities is to use a special form of maximum length sequence known as an Inverse Repeat PRBS (82). This sequence is formed, by inverting alternate outputs of the PRBS resulting in a sequence of twice the maximum length. The autocorrelation function of this sequence is depicted by fig 2.2.6. The study due to Godfrey and Moore (43) shows that for a first order process with time constant that may take one of two values depending on whether the output variable is increasing or reducing, then the output can be described as a simple polynomial function of the past inputs. The effect of the inverse repeat PRBS is to cancel the second order terms in the polynomial, and if

the higher order terms are negligible then the effect of the nonlinearity is removed. For the same conditions as fig 2.2.4 except that the input sequence is an inverse repeat PRBS, the crosscorrelation was obtained and is depicted by fig 2.2.7. The secondary peaks have been removed and the estimates in the settled portion are randomly disposed about zero, hence it may be assumed that the nonlinear effects have been reduced substantially.

The drift correction and nonlinearity compensation will be used with the crosscorrelation techniques described to obtain system model which will at this stage be assumed to be accurate. As with any practical modelling exercise there will always be a finite error (in some sense) between the model and the actual system, however the models derived will be shown, in section 2.3, to be adequate for design of controllers to meet certain performance specifications.

2.2.4 Parameter model identification using least squares techniques

This section presents techniques for identifying models of the form,

$$G(Z^{-1}) = \frac{B(Z^{-1})}{A(Z^{-1})}$$

where

$$B(Z^{-1}) = b_0 + b_1.Z^{-1} + b_2.Z^{-2} + ... b_n.Z^{-n}$$

and $$A(Z^{-1}) = 1 + a_1.Z^{-1} + a_2.Z^{-2} + ... a_n.Z^{-n}$$

This transfer function representation is equivalent to the weighting sequence but requires less parameters and is more convenient for control design.

The identification technique is based on a least squares algorithm (52) and is developed such that estimates are obtained as time evolves allowing changing parameters to be 'tracked'. This technique was chosen because it is used extensively in the self tuning control design, hence its use as a plant identifier allowed initial 'debugging' of the routines and a familiarity to be established with the technique and its limitations.

It is well known that the least squares parameter estimates will be biased in most cases if the plant is subjected to random disturbances, and the theory of generalised least squares (GLS) (12) is described in section 2.2.5 as a means of obtaining more accurate estimates.

The system can be described by the linear discrete model shown by fig 2.2.8, which is sufficiently general to satisfy the model structure described in sections 2.1.6 and 2.1.7. The system equation for this representation is given by the following:

$$y(t)=Z^{-k}.B(Z^{-1}).U(t)+(1-A(Z^{-1})).y(t)+\frac{A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}.\xi(t)+A(1).d \qquad (2.2.12)$$

where '$\xi(t)$' is a member of an uncorrelated random sequence and '$C(Z^{-1})/D(Z^{-1})$' is a stable and inverse stable transfer function.

The terms 'd' and '$\xi(t)$' account for random disturbances, constant errors, and measurement errors. In practice due to ambient variations the noise process will appear to be nonstationary. This effect is included in the model by making the disturbance 'd' slowly time varying.

Defining the residual '$\varepsilon(t)$' as the combined disturbance terms as follows,

$$\varepsilon(t) = \frac{A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}.\xi(t) + A(1).d \qquad (2.2.13)$$

Then the system equation can be expressed in the standard form for regression analysis as follows,

$$y(t) = \underline{x}^T(t).\underline{\theta} + \varepsilon(t) \qquad (2.2.14)$$

where

$$\underline{x}^T(t) = (U(t-k), U(t-k-1), \ldots U(t-k-n), y(t-1), y(t-2), \ldots y(t-n))$$

$$\underline{\theta}^T = (b_0, b_1, \ldots b_n, -a_0, -a_1, \ldots -a_n)$$

The least squares identification algorithm estimates the regression coefficients '$\underline{\theta}$' to achieve a best fit to experimental data in the sense of minimum-error-squares. That is, the least squares estimate of the regression coefficients, '$\hat{\underline{\theta}}$', is estimated so that the following error function is minimised.

$$J = \sum_{t=n+1}^{N+n} \varepsilon^2(t) \Big|_{\theta = \hat{\theta}}$$

As this technique is developed for control purposes then it is desirable for the estimated parameters to 'track' any slow changes in the actual parameters. This is achieved by the weighting of past data so that more recent data has more effect on the estimates. For exponential weighting of past data the error function is given by,

$$J = \sum_{t=n+1}^{N+n} \beta^{N-t}.\varepsilon^2(t) \qquad (2.2.15)$$

$$0<\beta<1$$

The weighting factor 'β' is known as the 'forgetting factor', it effectively indicates the number of data points used by the estimator. The number of data points is given by the following relation.

$$\propto \cong \frac{1}{1-\beta}$$

The least squares estimates of the regression coefficients is given by the well known real-time recursive algorithm described by the following set of equations.

$$\begin{aligned} \underline{\hat{\theta}}(t) &= \underline{\hat{\theta}}(t-1)+\underline{K}(t).(y(t)-\underline{x}^T(t).\underline{\hat{\theta}}(t-1)) \\ \underline{K}(t) &= P(t).\underline{x}(t)/(\beta+\underline{x}^T(t).P(t).\underline{x}(t)) \qquad (2.2.16) \\ P(t+1) &= (I-\underline{K}(t).\underline{x}^T(t)).P(t)/\beta \end{aligned}$$

The regression coefficients are equal to the previous coefficient estimates plus the product of the fitting error and a vector, 'K(t)', known as the Kalman gain vector. As the equations update the regression coefficient vector '$\underline{\hat{\theta}}(t)$' and covariance matrix 'P(t)' recursively

then it is necessary to initially set these at the start up of the algorithm. The initial setting of the regression coefficients will determine to a significant extent the speed of convergence of the estimates, and as will be shown later, the convergence to the correct values of GLS estimates is determined largely by the initial estimates. The initial setting of the covariance matrix is usually given by the following relation,

$$P(0) = f.I.$$

where 'I' is the identity matrix and 'f' is a large positive scalar. For 'f' large enough (1000 say) then the recursive estimates are asymptotically the same as the least squares estimates.

The basic algorithm as given by equation 2.2.16 is susceptible to rounding errors, this phenomenon being quickly realised for short word length number representations. To overcome the ill conditioning due to rounding errors the estimation algorithm can be modified so that the square root of the covariance matrix is modified rather than the matrix itself. Such an algorithm, presented by Clarke et al (15) based on the results of Peterka (71), is used to obtain numerically stable parameter estimates.

It is now convenient to discuss the statistical properties of the regression coefficient estimates. The choice of input signal as a PRBS ensures that a solution exists for '$\underline{\hat{\theta}}$', as in the case of crosscorrelation this is an optimum test signal. To examine the properties of this solution

it is necessary to examine the statistics of the residual '$\varepsilon(t)$'. If it is assumed that the residual is a member of an uncorrelated zero mean random sequence which is independent of the data vector '$\underline{x}(t)$', then the estimates are unbiased, that is

$$E(\hat{\theta}) = \theta$$

and consistent, that is

$$\underset{N \to \infty}{\text{Limit}} \ \hat{\theta} = \theta$$

However, the residual is in fact given by equation 2.2.13 which is nonzero mean and correlated. Hence in general the least squares estimates will be biased.

2.2.5 Generalised least squares techniques

In an effort to reduce the bias in least squares estimates due to correlated residuals, the method of generalised least square (GLS) was developed (12). The method due to Clarke (12) is based on the prefiltering of the data vector so that the residual becomes uncorrelated or 'white'. This prefilter is commonly known as the 'whitening' filter. This section describes ways of estimating the noise process for prefiltering and techniques for direct removal of bias in estimates.

The data vector must, in general, comprise of zero mean data, and hence the inclusion of the constant term 'd' in equation 2.2.13 may account for any error in calculation of the perturbation. The data must be of zero mean, because the large signal gains will be

different from small signal gains, and if the perturbation is insignificant with respect to the mean value then the steady state gains will be identified rather than the dynamics. Mean values are likely to vary due to the nonstationary nature of the disturbances on the plant, hence estimation of mean values is usually carried out recursively with exponential weighting of data (54). Calculation of the perturbation is given simply by subtraction of the estimated mean value from the signal value. This is equivalent to prefiltering the data by the following filter.

$$Pl(Z^{-1}) = \frac{(1-Z^{-1}).\lambda}{(1-\lambda.Z^{-1})}$$

where 'λ', the forgetting factor, is in the range

$$0 < \lambda < 1$$

The data length for which the average is calculated is given by the relation

$$N = \frac{1}{1-\lambda} \text{ samples}$$

An alternative method of removing constant errors is to difference the data, that is prefilter by the following filter.

$$Pl(Z^{-1}) = (1-Z^{-1})$$

For either prefilter the residual given by equation 2.2.13 becomes,

$$\varepsilon(t) = \frac{Pl(Z^{-1}).A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}.\ \xi(t)+Pl(1).A(1).d$$

and as,

$$Pl(1) = 0$$

then,

$$\varepsilon(t) = \frac{Pl(Z^{-1}).A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}.\xi(t) \qquad (2.2.17)$$

Therefore the residual is zero mean, and correlated.

For sufficiently high signal to noise ratio the high pass filtering described is sufficient to obtain accurate estimates, however for reduced signal to noise ratio the residuals must be whitened to obtain unbiased estimates. To prefilter it is necessary to have an estimate of the noise process, and in the estimation techniques to be presented it is assumed that the noise process can be sufficiently accurately represented by the autoregression given by;

$$\varepsilon(t) = \frac{\xi(t)}{F(Z^{-1})} \qquad (2.2.18)$$

where

$$F(Z^{-1}) = 1+f_1.Z^{-1}+f_2.Z^{-2}+ \ldots. f_pZ^{-p}$$

Therefore the regression equation can be expressed in the following forms

$$F(Z^{-1}).A(Z^{-1}).Y(t)=F(Z^{-1}).B(Z^{-1}).U(t)+\xi(t) \qquad (2.2.19)$$

and

$$A(Z^{-1}).Y(t)=B(Z^{-1}).U(t)+(1-F(Z^{-1})).\varepsilon(t) +\xi(t) \qquad (2.2.20)$$

The first form shows that if the noise autoregression is known then the data vector may be filtered such that

$$y^*(t) = y(t).F(Z^{-1}),\ U^*(t) = U(t).F(Z^{-1})$$

The regression equation is then given by,

$$A(Z^{-1}).Y^*(t) = B(Z^{-1}).U^*(t) + \xi(t)$$

and as the noise term '$\xi(t)$' is zero mean and uncorrelated then least squares techniques can be applied to obtain unbiased estimates of '$A(Z^{-1})$, $B(Z^{-1})$'. This is the basis of the study due to Clarke (12) and will be described later.

The second form in equation 2.2.20 suggests that by extending the regression coefficient vector such that it is given by,

$$\underline{\theta}^T = (b_0, b_1, \ \ldots\ b_n, -a_1, -a_2, \ \ldots\ -a_n, -f_1, -f_2, \ \ldots\ -f_p)$$

then the noise autoregression may be estimated along with the plant. This is the basis of the study due to Hsia (51) and will be described presently.

All GLS techniques are based on the assumption that the error function 'J', which is obtained from equations 2.2.15, 2.2.19 and is given by,

$$J = \Sigma(F(Z^{-1}).A(Z^{-1}).y(t) - F(Z^{-1}).B(Z^{-1}).U(t))^2$$

can be minimised with respect to A, B and F independently. However as the cost function is a nonlinear function of the

unknown parameters then minimisation can only be approximate and hence there are no general convergence proofs for estimates from GLS procedures. Because of the necessary numerical nonlinear minimisation procedure, several different GLS techniques have been proposed (12, 51, 84, 3). The repeated least squares technique due to Astrom (3), which is not strictly of GLS type, will not be investigated because of the necessary large number of parameters to estimate and complex manipulations that are required.

The techniques of Hsia (51) and Talmon and van den Boom (84) will now be described. These techniques are attractive because prefiltering is unnecessary as bias is removed from the estimates directly. The technique due to Hsia will be known as extended least squares (ELS). From equation 2.2.20 the estimated regression coefficient vector is given by.

$$\underline{\hat{\theta}}^T = (\hat{b}_0, \hat{b}_1, \ldots \hat{b}_n, -\hat{a}_1, -\hat{a}_2, \ldots -\hat{a}_n, -\hat{f}_1, -\hat{f}_2, \ldots -\hat{f}_p)$$

and the data vector

$$\underline{x}^T(t) = (U(t), U(t-1), \ldots U(t-n), Y(t-1), Y(t-2), \ldots Y(t-n), \hat{\varepsilon}(t-1), \hat{\varepsilon}(t-2), \ldots \hat{\varepsilon}(t-p))$$

where '$\hat{\varepsilon}(t)$' is the estimated residual obtained from the estimated regression equation given by,

$$\hat{A}(Z^{-1}).Y(t) = \hat{B}(Z^{-1}).U(t) + \hat{\varepsilon}(t)$$

Therefore the estimated residual is generated by a time varying system, the variation of which is dependent on the initial estimates of the regression coefficients. Although there are no convergence proofs for this algorithm it is known that the convergence to the correct estimates is dependent on the initial choice of regression coefficients '$\hat{A}(Z^{-1})$, $\hat{B}(Z^{-1})$'. An estimate of these parameters can be obtained from the biased least squares estimates, however the adequacy of these estimates is dependent on the signal to noise ratio. Alternatively the initial estimates may be obtained from the results of a step response test and this was shown to lead to satisfactory convergence. This technique will be used in later experimental studies.

The technique due to Talmon and van den Boom (84) is an extension to the technique previously described, that is the noise process is described by

$$\varepsilon(t) = \frac{Fn(Z^{-1})}{Fd(Z^{-1})}.\xi(t)$$

which gives

$$\varepsilon(t)=(fn_1.Z^{-1}+fn_2.Z^{-2}+\ldots fn_q Z^{-q}).\xi(t)-(fd_1.Z^{-1}+ fd_2.Z^{-2}+ \ldots. fd_p.Z^{-p}).\varepsilon(t)+\xi(t)., \quad fd_o=1.$$

The estimated regression coefficient vector and data vector are given by,

$$\underline{\hat{\theta}}^T=(\hat{b}_0,\hat{b}_1,\ldots\hat{b}_n,-\hat{a}_1,-\hat{a}_2,\ldots-\hat{a}_n,-\hat{fd}_1,-\hat{fd}_2\ldots-\hat{fd}_p, \hat{fn}_1,\hat{fn}_2,\ldots\hat{fn}_q)$$

$$\underline{x}^T(t)=(U(t),U(t-1),\ldots U(t-n),y(t-1),y(t-2),\ldots$$
$$Y(t-n),\hat{\varepsilon}(t-1),\hat{\varepsilon}(t-2)\ldots\hat{\varepsilon}(t-p),\hat{\xi}(t-1),\hat{\xi}(t-2),$$
$$\ldots.\hat{\xi}(t-q))$$

where the estimated residual is given by,

$$\hat{\varepsilon}(t)=\hat{A}(Z^{-1}).Y(t)-\hat{B}(Z^{-1}).U(t)$$

and estimated noise variable '$\hat{\xi}(t)$' is given by,

$$\hat{\xi}(t)=\hat{Fd}(Z^{-}).\hat{\varepsilon}(t)+(1-\hat{Fn}(Z^{-1})).\hat{\xi}(t)$$

As with the previous technique no convergence proofs exist for estimated parameters. Furthermore the convergence to the correct values is not only dependent on the initial estimates of the plant parameters, '$A(Z^{-1}),B(Z^{-1})$', but also on the initial estimates of the noise parameters, '$Fd(Z^{-1})$, $Fn(Z^{-1})$'. Initial estimates of noise parameters are not easily available from simple tests, and as a result convergence is poor. Hence this technique will not be further developed.

The GLS technique due to Clarke (12) was developed primarily for offline processing, however a simplified version has been extended to on line computation (49). The Clarke method requires the repeated processing of a large data record of 'N' data points, where 'N' may be 1000. The noise autoregression is expressed as the product of a number of cascaded autoregressive filters, such that

$$F(Z^{-1}) = \prod_{i=1}^{m} \{1+F_i(Z^{-1})\}$$

The individual cascaded filters are identified from the residual sequence generated at each processing of the data record. As each filter is estimated the data prefilter is multiplied by the estimate. The plant parameters are then estimated using the prefiltered data. This process is repeated until the value of 'F_i' tends to zero and hence the cascade tends to a fixed value.

This method has the advantage that a low order filter can be identified at each stage, however the disadvantage is that a large number of iterations may be necessary due to the slow convergence. This method may also converge to the wrong answer (83). The technique used in later sections is a simplification of this method, in that the prefilter is identified in total at each stage, thus removing the necessity of forming the prefilter from a cascade of identified filters. This technique will be referred to as GLS.

All the techniques described require that the noise process is expressed by the autoregression of equation 2.2.18 therefore it is important to consider the practicality and accuracy of such an assumption. The noise autoregression, from equations 2.2.17, 2.2.18 is given by the following relation.

$$\frac{1}{F(Z^{-1})} = \frac{Pl(Z^{-1}).A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}$$

If the high pass filter '$Pl(Z^{-1})$' has a zero on the unit circle then unless '$D(Z^{-1})$' has a similar zero then the noise autoregression must approximate this

factor. The equivalent autoregression to a zero on the unit circle is given by,

$$(1-Z^{-1}) = \frac{1}{1 + \sum_{i=1}^{\infty} Z^{-i}}$$

The relation is a main reason for high order noise autoregressions to be necessary when the signal to noise ratio is low. If '$D(Z^{-1})$' does have a zero on the unit circle, which is unlikely, then the disturbance is the special form of noise process known as Brownian motion. As will be demonstrated by experiment the resultant high order noise process is a major problem when signal to noise ratio is low.

In all the GLS techniques described a reduction in the order of the autoregressive prefilter was shown to be possible by the expression of the residual in the following form.

$$\varepsilon(t) = \frac{A(Z^{-1})}{F(Z^{-1})} \xi t$$

in which case the noise process is given by,

$$\frac{A(Z^{-1})}{F(Z^{-1})} = \frac{Pl(Z^{-1}).A(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}$$

therefore

$$\frac{1}{F(Z^{-1})} = \frac{Pl(Z^{-1}).C(Z^{-1})}{D(Z^{-1})}$$

The prefiltered output and input are given by,

$$Y^*(t) = \frac{\hat{F}(Z^{-1})}{\hat{A}(Z^{-1})}.Y(t), \quad U^*(t) = \frac{\hat{F}(Z^{-1})}{\hat{A}(Z^{-1})}.U(t)$$

where '$\hat{A}(Z^{-1})$' is obtained from the regression coefficient vector and '$\hat{F}(Z^{-1})$' by processing of the residuals as described.

Prefiltering by '$\hat{A}(Z^{-1})$' was also found to increase the speed of convergence for both GLS and ELS, and in the latter case the prefilter was updated for each processing of the data record in the same manner as for GLS.

2.3 Identification of the heating system

As previously stated, the object of the identification studies is to estimate 2nd Order discrete models of the following form,

$$G^*(Z^{-1}) = \frac{Z^{-k}.(b_1.Z^{-1}+b_2.Z^{-2})}{(1+a_1.Z^{-1}+a_2.Z^{-2})} \qquad (2.3.1)$$

that adequately describe the dynamics of heating systems. Equation 2.3.1 is the discrete equivalent of the following 2nd order continuous time transfer function with zero order hold.

$$G(S) = \frac{K.\exp(S.D)}{(S+a)(S+b)}$$

where

$$D = (k-1).ST$$

and 'k' is a positive integer.
However, in practice the time delay is not an integer multiple of the sample time and a more realistic discrete transfer function is the following

$$G(Z^{-1}) = Z^{-k}\frac{(b_0+b_1.Z^{-1}+b_2.Z^{-2})}{(1+a_1.Z^{-1}+a_2.Z^{-2})} \qquad (2.3.2)$$

In this section the two types of model, equations 2.3.1, 2.3.2, will be identified and the results compared with respect to accuracy of representation of the actual system response. Obviously it would be advantageous to use the form of equation 2.3.1 for control design since it requires one less parameter and hence identification is eased.

In all the experiments to be described transfer functions are identified using zero mean data, this being achieved by prefiltering input/output data by the following differencing filter.

$$Pd(Z^{-1}) = (1-Z^{-1})$$

The pure time delay term 'k' is a useful parameter when modelling high order systems with low order models, and there can be several methods of choosing the value of 'k' depending on the model fitting criteria chosen. For these experiments it was considered sufficient to choose 'k' equal to the number of delays of the first non zero weighting sequence estimate. This value turned out to be unity because of the small transport lag of the heating medium and the lack of small time constant heat stores within the duct and space.

The disturbances on the test facility are due to ambient variations and electrically coupled measurement noise. Since the ambient conditions vary slowly with respect to the time constants of the system then the differencing filter will remove the majority of this disturbance from the data.

The effects of the measurement noise can be reduced to a minimum by low pass analogue filtering of the transducer inputs, however the disturbances cannot be completely removed and so LS estimates will be biased. At this stage no assumptions are made as to the exact nature of the disturbances on the test facility, however for one of the experiments the measurements were corrupted artificially by generated white noise. The noise is formed by the addition of twelve uniformly distributed random variables and so the resulting distribution is approximately Gaussian (15).

So that the adequacy of the derived transfer function for the heat exchanger (equation 2.2.1) may be examined and to initially develop the identification routines, the duct temperature is used as the output and the results described in section 2.3.1. Section 2.3.2 describes the experiments using exhaust temperature as system output to obtain an overall model for the environmental test facility including heat exchanger and transducers.

Since the final objective is control design based on identified models, then the modelling accuracy is a

function of how closely the control performance is to that predicted. As an example of how model accuracy effects predicted control performance, a closed loop control is applied to regulate exhaust temperature and a comparison made of predicted and measured control performance criteria. The performance criteria were chosen as the variance or mean squared value of temperature with respect to set value and the variance of control output. So that noise characteristics were known the measurements are corrupted by white noise of sufficient magnitude to swamp the natural noise and disturbances. The control law investigated is proportional plus integral (P+I) and the results are presented in chapter 2.3.3.

2.3.1 Identification of the heat exchanger transfer function

The heat exchanger transfer function will be defined as the description of the dynamics relating power demand signal and duct temperature monitored immediately after the heat exchanger. Due to the rapid dynamic response of the heat exchange process, the dynamics of the temperature transducer will contribute a significant part to the overall response.

To obtain the weighting sequence the duct temperatures and power demand signal were crosscorrelated for the following experimental conditions,

input signal = 50% ± 10%, 63 stage PRBS, 9 sequences applied

sample period = 10 seconds

and the resulting weighting sequence is given in fig 2.3.1. The identification of the transfer function is significantly simplified as a result of a high signal/noise power ratio. This is due partly to the low losses within the duct and also due to radiant heat exchange between heater and detector surfaces. The transfer function was identified assuming only a moving average noise process given by,

$$\varepsilon(t) = A(Z^{-1}).\xi(t)$$

in which case the data whitening filter is of the form

$$Pw(Z^{-1}) = \frac{1}{\hat{A}(Z^{-1})}$$

and close correlation was obtained between weighting sequence and estimated transfer function pulse response. The result of three iterations of the whitening filter, for 1000 samples of data, and for the two types of model gave the following results

$$G^*(Z^{-1}) = \frac{0.0094.Z^{-1}+0.02.Z^{-2}}{1-1.274.Z^{-1}+0.364.Z^{-2}}$$

$$G(Z^{-1}) = \frac{0.0095.Z^{-1}+0.022.Z^{-2}+0.0076.Z^{-3}}{1-1.03.Z^{-1}+0.13.Z^{-2}}$$

The initial estimates for the iteration process were taken as the LS estimates given by

$$G\ell s(Z^{-1}) = \frac{0.006.Z^{-1}+0.016.Z^{-2}}{1-1.02.Z^{-1}+0.27.z^{-2}}$$

The pulse responses of these three transfer functions are superimposed on the weighting sequence to compare accuracy

as in fig 2.3.1.a, fig 2.3.1.b, fig 2.3.1.c.

As a moving average noise process was assumed then a comparison between GLS and ELS techniques could not be carried out.

2.3.2 Identification of the test facility 'lumped parameter' model

The lumped parameter model is defined as the description of the dynamics relating exhaust temperature to power demand signal.

To obtain the weighting sequence the exhaust temperature and power demand signal were crosscorrelated for the following experimental conditions: input signal = 50% ± 12%, 63 stage PRBS, 19 sequence applied.

sample period = 40 seconds

and the resulting weighting sequence is given in fig 2.3.2. Due to the reduced signal/noise power ratio it was necessary to assume a relatively complex ARMA noise process of the form,

$$\varepsilon(t) = \frac{A(Z^{-1})}{F(Z^{-1})}.\xi(t)$$

and for satisfactory convergence '$F(Z^{-1})$' is of 6th order. Further, the low signal/noise value caused large bias in the LS estimates, given by,

$$G\ell s(Z^{-1}) = - \frac{0.00044.Z^{-1} + 0.0035.Z^{-2} + 0.00395.Z^{-3}}{1-0.108.Z^{-1} - 0.27.Z^{-2}}$$

such that these estimates were so much in error as to cause slow convergence for GLS and ELS. Instead the step response Z transform obtained from fig 2.2.2 was used as initial estimates and is given by,

$$G_{sr}\ (Z^{-1}) = \frac{0.0079\ Z^{-2}}{1-0.9.Z^{-1}} \qquad (2.3.3)$$

The order of the noise autoregression was chosen as a compromise between slow convergence due to low order or slow computation for high order.

The estimates for the GLS algorithm converged after 6 iterations of the ARMA noise filter for 500 samples of data, and the transfer function estimate is given by

$$Gg\ell s(Z^{-1}) = \frac{0.00092.Z^{-1} + 0.006.Z^{-2} + 0.0028.Z^{-3}}{1-1.175.Z^{-1} + 0.296Z^{-2}} \qquad (2.3.4)$$

The mean squared residual is given by

$$E\{\varepsilon^2(t)\} = 0.0551$$

The estimates for the ELS algorithm converged after 8 iterations of the moving average noise filter for 500 samples of data, the transfer function estimate is given by

$$Ge\ell s(Z^{-1}) = \frac{0.0012.Z^{-1} + 0.0058.Z^{-2} + 0.0042.Z^{-3}}{1-1.015.Z^{-1} + 0.135.Z^{-2}} \qquad (2.3.5)$$

The mean squared residual is given by

$$E\{\varepsilon^2(t)\} = 0.0448$$

The pulse responses for equations 2.3.4, 2.3.5 are superimposed on the weighting sequence in fig 2.3.2.a, fig 2.3.2.b. The slight error between pulse response and weighting sequence is not reduced by increase of order of plant or noise transfer function, hence it may be assumed this is the best linear estimate obtainable using these techniques.

The effect of noise magnitude and type of assumed noise filter on the accuracy of the estimates obtaining using the GLS algorithm can be demonstrated by increasing the measurement noise artificially. The additive noise is white and of variance 0.25. Transfer functions were estimated from the same data used in the previous experiment for several different assumed noise processes. The estimated transfer functions and mean squared residuals are given by,

$$G1(Z^{-1}) = \frac{-0.00064.Z^{-1}+0.0031.Z^{-2}+0.0055.Z^{-3}}{1-1.083.Z^{-1}+0.448.Z^{-2}}$$

$$E\{\varepsilon^2(t)\} = 0.0654$$

where assumed noise process is 6th order autoregressive

$$G2(Z^{-1}) = \frac{-0.572x10^{-5}.Z^{-1}+0.00483.Z^{-2}+0.0056.Z^{-3}}{1-1.168.Z^{-1}+0.35.Z^{-2}}$$

$$E\{\varepsilon^2(t)\} = 0.0629$$

where the assumed noise process is 12th order autoregressive

$$G3(Z^{-1}) = \frac{0.0007.Z^{-1}+0.0048.Z^{-2}+0.0053.Z^{-3}}{1-1.169.Z^{-1}+0.311.Z^{-2}}$$

$$E\{\varepsilon^2(t)\}= 0.0606$$

where assumed noise process is 12th order autoregressive and 2nd order moving average. The pulse responses from these transfer functions are plotted in fig 2.3.3.a, fig 2.3.3.b, fig 2.3.3.c., superimposed on the weighting sequence. These responses show that bias is reduced for increasing the order of the assumed autoregressive noise process and that defining the moving average noise process as '$\hat{A}(Z^{-1})$' further reduces bias and is easy to incorporate in the algorithm.

A reduced parameter model is identified using the GLS algorithm which is given by

$$G(Z^{-1}) = \frac{0.00071.Z^{-1}+0.007.Z^{-2}}{1-1.337.Z^{-1}+0.44.Z^{-2}} \qquad (2.3.6)$$

The pulse response is plotted in fig 2.3.4.a with the weighting sequence and the pulse response of the transfer function given by equation 2.3.4, is plotted against weighting sequence in fig 2.3.4.b.

2.3.3 Comparison of predicted and measured control performance

The control performance is defined by the variance of the plant input and output. The output variance is important for quality regulation and input variance gives an indication of the wear on the final control element, which in practice

would be a steam or water valve.

The closed loop system for unity feedback and for ward path controller can be depicted as fig 2.3.5, this can be modified to the regulation model, fig 2.3.6, if all temperatures are referred to a constant set value. The regulation model is a realistic linear model since set value changes occur infrequently (every 12 hours say) and when they do a nonlinear form of control is usually used to obtain a fast transient with little overshoot (55). This type of control may comprise a switched PI control, such that for large error the control is purely 'P' and as error reduces below a certain value then the 'I' term is switched in. This form of controller has been described by Krikelis (58) and termed an 'intelligent integrator'.

The system output 'Y(t)' can be assumed to be of zero mean as the integral term will compensate for the disturbance 'd(t)' which is due to the slow ambient changes. The input 'U(t)' is non zero mean due to the disturbance 'd(t)'. However, if the integral term is included in the plant transfer function then the control output will be zero mean, and hence the variance can be calculated.

For disturbance 'd(t)' constant or slowly time varying the system input and output can be expressed in terms of the noise variable 'ξ(t)' as the following.

$$Y(t) = \frac{\xi(t).C(Z^{-1})/D(Z^{-1})}{1+Z^{-k}.B(Z^{-1}).H(Z^{-1})/A'(Z^{-1})}$$

and

$$U(t) = \frac{-\xi(t).C(Z^{-1}).H(Z^{-1})/D(Z^{-1})}{1+Z^{-k}.B(Z^{-1}).H(Z^{-1})/A'(Z^{-1})}$$

where

$$A'(Z^{-1}) = (1-Z^{-1}).A(Z^{-1})$$

that is, the integral term (for rectangular integration) is included in the plant denominator. In which case the control is given by the following.

$$H(Z^{-1}) = h_0 + h_1.Z^{-1}$$

The measurement noise is chosen, for simplicity, to be uncorrelated, in which case the noise process is

$$\frac{C(Z^{-1})}{D(Z^{-1})} = 1$$

The variance of input and output are given by the following (2),

$$E\{Y(t)^{2}\} \cong E\{\xi(t)^{2}\}.\sum_{i=1}^{N1} (W1_i)^2$$

and

$$E\{(U(t)^{2}\} \cong E\{\xi(t)^{2}\}.\sum_{i=1}^{N2} (W2_i)^2$$

where

$$\sum_{i=1}^{N1} W1_i.Z^{-i} \cong \frac{1}{1+Z^{-k}.B(Z^{-1}).H(Z^{-1})/A'(Z^{-1})}$$

and

$$\sum_{i=1}^{N2} W2_i.Z^{-i} \cong \frac{H(Z^{-1})}{1+Z^{-k}.B(Z^{-1}).H(Z^{-1})/A'(Z^{-1})}$$

These variances will be calculated and compared to the measured values from practical experiment.

So that the variances can be measured the exhaust temperature is regulated by a control given by,

$$U(t) = (17-15.Z^{-1}).(W(t)-TX(t))$$

where 'W(t)' the set value is given by

$$W(t) = TA(t)+3^{o}C$$

As the ambient temperature varies slowly then this variable setvalue will not violate the regulation assumption. The slowly time variable set value was necessary since the restricted control range means that the exhaust temperature can only be raised $6^{o}C$ above ambient at maximum. Therefore fixing the setvalue midway in this range ensured that the control average value is also midway in its range. This then reduces the chances of the control entering nonlinear regions at the extreme ends of its range if the ambient temperature drifts more than $3^{o}C$ during the experiment.

The measured output is given by

$$Y(t) = TX(t) + \xi(t)$$

where '$\xi(t)$' is white noise of variance 0.25.

The control was carried out over a 24 hour period and the variance measured after the initial transient had

died away. The variances are given by

$$E\{(U(t))^2\} = 135.5$$

$$E\{(Y(t))^2\} = 0.289$$

For zero control the variances are given by

$$E\{(U(t))^2\} = 0$$

$$E\{(Y(t))^2\} = 0.25$$

The variances will be predicted from three models, which are of varying accuracy with respect to the weighting sequence. The models and variances are given by

step response model equation 2.3.3

$$E\{(U(t))^2\} = 130.5 \qquad E\{(Y(t))^2\} = 0.2735$$

GLS reduced parameter model equation 2.3.6

$$E\{(U(t))^2\} = 131.68 \qquad E\{(Y(t))^2\} = 0.2793$$

ELS model equation 2.3.5

$$E\{(U(t))^2\} = 132.63 \qquad E\{(Y(t))^2\} = 0.2835$$

The model that gives the closest correspondence to the weighting sequence, the ELS model, also gives the closest prediction of control performance. However, the improvement over the step response model is only slight, due to the insensitivity of the performance criteria to modelling errors and the accuracy of the step response model. The accuracy of the first order model pulse response can be seen by comparing it to the reduced parameter model pulse response, see fig 2.3.4.c.

2.4 Conclusion of modelling and identification studies

The object of the modelling and identification study is to obtain a sufficiently accurate system description so that a controller can be designed. In lieu of a control design specification on model accuracy a judgement has been made on accuracy by comparing the weighting sequence and the pulse response of the estimated transfer functions.

In section 2.1.1 it was stated that the dynamics of a heat exchanger of an arbitrary type may be adequately described by a 2nd order transfer function in the form of equation 2.1.1, the discrete equivalent is given by equation 2.3.3 or 2.3.1. The heat exchanger transfer function is defined as the transfer function relating duct temperature to power demand signal. This transfer function was identified and the pulse response compared to the weighting sequence, see fig 2.3.1.a, fig 2.3.1.b., fig 2.3.1.c. The degree of correlation between the weighting sequence and pulse response is excellent thus justifying the assumption of a 2nd order model. The reduced parameter transfer function (equation 2.3.1) gave slightly less accurate correlation but for control design would be perfectly adequate.

The 2nd order transfer function relating exhaust temperatures to power demand signal was identified using the GLS and ELS algorithms, and the results are compared to the weighting sequence in fig 2.3.2.a, fig 2.3.2.b. The correlation between weighting sequence and pulse

response is not as close as for duct temperature (see fig 2.3.1.a,fig 2.3.1.b., fig 2.3.1.c., this being primarily attributable to the effects of dynamic nonlinearities such as that caused by imperfect mixing of air in the space. These nonlinearities are also indicated by the secondary peaks in the weighting sequence at 51, 55, 60 samples. The major secondary peak is at 60 samples which complies well with that predicted by Godfrey and Moore (43). The errors are not attributed to underestimation of order of plant or noise transfer functions since increasing the order does not reduce the error.

Inspection of fig 2.3.4.a., fig 2.3.4.b., shows that the difference in accuracy of representation between the full and reduced parameter 2nd order model is insignificant Thus stengthening the case for use of simple 2nd order models to describe complex processes of high order.

For increased noise magnitude it has been shown that the noise process must be more accurately modelled to reduce bias to acceptable levels (see fig 2.3.3.a, fig 2.3.3.b, fig 2.3.3.c. If the noise process is approximated by an autoregression then its order may be very high (>10) if the noise process has zeroes on the unit circle. This situation arises when high pass filtering is introduced to remove non zero mean values. Rather than prefiltering data to remove DC values, a constant value may be estimated along with the unknown parameters. This scheme however led to very slow convergence, and for the noise

levels experienced, the process of differencing and prefiltering to whiten residuals gave more satisfactory results.

Inspection of fig 2.3.3.a, fig 2.3.3.b, fig 2.3.3.c, shows that the approximation of the noise process by an ARMA filter where the estimated plant transfer function denominator is the moving average part leads to reduced bias. This is an attractive scheme since the increased prefiltering is negligible and the reduction in estimated parameters is significant. There may however be situations where this scheme is not applicable, in particular where the noise process of equation 2.2.17 is given by,

$$\varepsilon(t) = Pl(Z^{-1}).C(Z^{-1}).\xi(t)$$

therefore

$$D(Z^{-1}) = A(Z^{-1})$$

The relative merits of GLS and ELS techniques may be compared with respect to accuracy, speed and convenience of estimation. In pure accuracy terms the ELS estimates are better since they appear to yield a smaller mean squared residual value than GLS estimates. However the improvement is so small as to be insignificant for control design purposes. The number of iterations of the whitening filter for ELS and GLS is 8 and 6, and the time taken for one iteration is 120 and 100 seconds respectively. Thus the GLS algorithm is more complex than

than ELS firstly because the data must be processed twice for each iteration, once to estimate plant parameters and once to estimate noise parameters, and secondly due to the prefiltering operation.

In section 2.3.3 it was shown how the closed loop performance criteria relate to modelling accuracy, in particular how the model that yields the smallest modelling error also can be used to predict the closed loop performance most accurately. Although the identified model predicts the performance accurately it is only marginally more accurate than the simple model obtained from the step response. This can be attributed to the insensitivity of the performance criteria to modelling errors and also to the fact that the simple model is a 'good' model of the system. The simple first order model is a 'good' model of the system because the air space does not contain any material that would act as a significant heat store, other than the air, thus the temperature response is first order dominated.

3 Design of controls for HVAC systems

The layout of a typical heating, ventilating, and air conditioning (HVAC) system has been described in section 1.1, and the means by which the temperature and humidity are varied have also been presented. In this chapter techniques will be described for the design of the control mechanisms that alter the input of heat, via the final control element into the air space, to achieve a given temperature setvalue. It is not the intention to investigate humidity control, however it is considered that the principles derived for temperature control are applicable also to humidity.

The identified model of the test facility (equation 2.3.6) does not indicate the presence of significant transport lag. This can be attributed to the very short distance between heat exchanger and space, the large number of air changes per hour, and the lack of heat stores in the path ot the supply air. It is proposed that these practical phenomena may be adequately modelled by the addition of pure time delay to the system. This can be simulated by delaying the control signal by an integer number of samples. Besides making the test facility more representative of the real plant, this will also make control a more exacting task and hence aid the comparison of different control strategies. The effect of this excess time delay will be considered in section 3.1, in particular the effect on "conventional" control performance. By conventional it is meant techniques used

throughout the process control industry at this present time. It can be argued that the savings obtained by correct tuning for process control loops is greater than for HVAC control loops and hence control timing is usually more accurate for process loops. However so that comparison can take place between conventional and less conventional control schemes it is thought wise to tune the conventional control to a relatively high standard.

A typical design procedure would start by choosing the form of the control algorithm, (PID, PI, P, nonlinear) this choice being based on engineering experience and analysis. In this case, linear PI control will be used because of excellent results reported for its use (77) when compared to techniques such as Smith predictor (64) control and other less rigorous schemes. The PI control is particularly common in the process control field due mainly to its robust nature, (with respect to plant changes) and easily understood mode of operation.

The second step in the design procedure would be to carry out an experiment to obtain the significant characteristics of the plant. The characteristics obtained from a step response for instance, such as time constant and gain, can then be used in conjunction with tuning charts to obtain rough estimates for the control parameters values. Tuning charts due to Fertik (33) are used in this particular case.

The last step in the design procedure is then the final tuning by hand, if necessary, to achieve the desired performance. A control is designed in section 5.1 for the test facility using these techniques.

It will be shown that the effect of time delay is to make the system response more sensitive to plant parameters changes and to reduce stability margins. As a result it becomes necessary to reduce controller gain to ensure stable control and hence results in a degradation of performance. The results of section 3.2 show that this degradation can be largely avoided if instead of feeding back the measured output a predicted future output is fedback. Prediction of exact future system output implies the following two assumptions are justified:

(i) Disturbances on the system and measurement noise are known beforehand or are negligible.

(ii) A model of the plant has been obtained.

Assumption (i) can be removed if the prediction is based on obtaining the best (in some sense) prediction in spite of disturbances and noise. In this case the least squares predictor (2) is investigated. This form of predictor is convenient because the prediction equations are in such a format as to allow estimation of the unknown parameters using a simple recursive least squares technique. This then conveniently avoids assumption (ii).

Throughout Chapter 3 it will be assumed that the plant parameters are known exactly, unknown parameters being considered in chapter 4.

3.1 Conventional control

The so called three term control or PID control is very popular in process control but to a lesser extent in HVAC systems. This reduced popularity is due mainly to the difficulty of tuning three control parameters which necessitates more highly skilled personnel to commission the controls, and the resultant increase in the commissioning cost. Furthermore, in certain applications the measured temperature in the space may be so heavily contaminated with noise that the derivative term may cause a degradation in the performance. For these reasons the most popular form of control for HVAC systems is PI or two term control.

As has been mentioned, the presence of time delay in plant can cause difficulties in control tuning and this has been recognised for several years (64). The excess phase lag introduced by the time delay reduces stability margins and makes the response more sensitive to plant parameter changes. This phenomenon can be demonstrated by way of a simple example.

The block diagram of a system consisting of a lag plus delay plant and proportional control is shown by fig 3.1.1. To

illustrate the effect of time delay a graph of plant gain against controllability ratio will be constructed showing the regions of stability and critically damped response.

$\frac{L}{T}$ = controllability ratio

K.G = plant gain

The plant characteristic equation is given by

$$(1+S.T) + K.G.\exp(-S.L) = 0 \qquad (3.1.1)$$

The boundary of stability occurs when 's=j.w', which upon substituting for 's' in equation 3.1.1 gives the magnitude and angle criterion.

$$\left|\frac{K.G.\exp(-j.W.L)}{(1+j.W.T.)}\right| = 1$$

$$\therefore K.G = \sqrt{1+(W.T)^2} \qquad (3.1.2)$$

and

$$\left|\frac{K.G.\exp(-j.W.L)}{(1+j.W.T)}\right| = \Pi$$

$$\therefore -W.L = \Pi + \mathrm{Tan}^{-1}W.T \qquad (3.1.3)$$

Substituting 'W' in equation 3.1.3 for 'W' in equation 3.1.2 gives the relation for K.G. in terms of L/T, which after some manipulation is given by

$$\frac{L}{T} = \left(\Pi - \mathrm{Cos}^{-1}\frac{1}{K.G}\right) / \sqrt{(K.G.)^2 - 1} \qquad (3.1.4)$$

This equation gives the stability boundary which is

plotted in fig 3.1.2. The critically damped response parameters are found by differentiating the characteristic equation and equating to zero that is,

$$\frac{d}{ds}\,(K.G + \exp(S.L).(1+S.T)) = 0$$

$$\therefore\ s = -\left(\frac{1}{L} + \frac{1}{T}\right) \qquad (3.1.5)$$

Substituting 's' from equation 3.1.5 into equation 3.1.1 gives the following,

$$K.G = \frac{T}{L}.\exp\left(-1-\frac{L}{T}\right) \qquad (3.1.6)$$

This relationship is plotted on fig 3.1.2, showing the regions of overdamped, underdamped, and unstable response. This show that for gain greater than unity the region of stability reduces very rapidly for increasing gain. Proceeding with this example, it is assumed that the object of a given design specification is to set the control gain 'K' so that the response for a given controllability ratio is critically damped. To illustrate the sensitivity of the setting of plant gain with respect to the controllability ratio differentiate equation 3.1.6 with respect to L/T. This leads to the equation given by,

$$\frac{dK.G}{d\ L/T} = \left(-\exp\left(-1-\frac{L}{T}\right).\left[\frac{(1+L/T)}{(L/T)^2}\right]\right) \qquad (3.1.7)$$

This relation plotted against 'L/T' is given by fig 3.1.3. This shows that for large L/T the rate of change of KG is

very small. Hence for large variation in L/T the gain is only allowed to vary by a very small amount, therefore making tuning difficult.

This then illustrates the destabilising effects of the time delay and its effect on tuning of a simple proportional control.

3.2 Predictive control

If the time delay for the plant is equal to an integer (k) number of samples then the k-step-ahead output is defined as the first output affected by the present control signal.

The form of predictive control to be presented is based on the simple idea that if the k-step-ahead output can be fedback then the control can be designed as if the delay is zero. This technique is adequately described by Smith's Principle (64), in which the control law is such as to externalise the time delay from the control loop. Two equivalent control loops are shown in fig 3.2.1, the only apparent difference is that the control law has been chosen so that the time delay is not within the loop. The control law '1/Q' is designed assuming that there is no delay present. The control law '1/Q*' can be found by comparing the following closed loop transfer functions,

$$\frac{G.Z^{-k}./Q}{1+G/Q} \equiv \frac{G.Z^{-k}/Q^*}{1+G.Z^{-k}/Q^*}$$

which after manipulation yields the following control law,

$$1/Q^* = \frac{1/Q}{1+G.(1-Z^{-K})/Q}$$

This can be represented in block diagram form as fig 3.2.2.

The successful application of this scheme relies on the following two assumptions:

(i) External disturbances and measurement noise are negligible or can be predicted.

(ii) An accurate model of the plant is available.

The principle roles of negative feedback is to compensate for unknown disturbances, and to reduce the sensitivity of the response to plant parameter variation, thus giving better setpoint following than for open loop control.

However if assumptions (i) and (ii) are justified then it appears that feedback is not necessary and the plant can be controlled as well in open loop as in closed loop. However, in general the disturbances are not known or neglibible and the plant model is only known approximately. If it is assumed that the plant model is known accurately then it is left to predict the output in the presence of noise and disturbances. If, as in chapter 2, the noise and disturbances can be considered to be samples of a correlated random sequence then it is possible to make a statistical prediction of the output.

The least squares predictor, derived in Appendix A, gives the prediction of the k-step-ahead output of the linear system as follows. The plant equation is given by,

$$A(Z^{-1}).Y(t)=Z^{-k}.B(Z^{-1}).U(t)+C(Z^{-1}).\xi(t)+A(Z^{-1}).d \qquad (3.2.1)$$

and the prediction is,

$$Y^{*}(t+k|t) = \frac{F(Z^{-1})}{C(Z^{-1})}.Y(t) + \frac{G'(Z^{-1})}{C(Z^{-1})} + \frac{\delta}{C(1)} \qquad (3.2.2)$$

where 'δ' is a constant, and the prediction error is given by,

$$\tilde{Y}(t+k|t) = E(Z^{-1}).\xi(t+k) \qquad (3.2.3)$$

The prediction error variance, which is a minimum, increases with delay 'k', and is given by,

$$E\{(\tilde{Y})^2\} = E\{(\xi(t))^2\}.(1+e_1^2+e_2^2+ \ldots . e_{k-1}) \qquad (3.2.4)$$

The control loop with least squares predictor is depicted by fig 3.2.3.

The least squares predictor is by definition a better predictor (in the least squares sense) than the Smith predictor but both require an accurate model of the plant. However the least squares predictor lends itself to inclusion in a particularly simple self tuning scheme which avoids the necessity of an accurate plant model. Such a scheme will be described in the next chapter.

4 Self tuning control

A self tuning controller consists of three main functional elements: the recursive parameter estimator, the control law design algorithm, and the control law, as in fig 4.1. The recursive parameter estimator uses plant input/output data to estimate the parameters of a discrete transfer function. This may be achieved by using techniques such as recursive least squares or recursive maximum likelihood identification. The control design algorithm calculates the coefficients of the control law based on the parameters output by the recursive parameter estimator. The control law, which is in the form of a difference equation, uses the desired set value, the fedback system output, and in certain cases feed forward signals to calculate the next control signal.

It is possible to significantly simplify the self tuning control algorithm by the omission of the control design algorithm. This being achieved by the reformulation of the input/output data fed to the parameter estimator, such that the parameter estimator produces estimates of the control law coefficients directly. This scheme is then known as an "implicit' self tuner, which distinguishes it from an "explicit" self tuner in which the plant parameters are estimated explicitly, the control coefficients being calculated based on these estimates.

The functions of parameter estimation and control signal calculation in a self tuner are separate operations, thus

the control signal is calculated assuming that the parameter estimates are equal to the actual values. This is known as the "certainty equivalence" (15) approach to controller design. This approach, although relatively simple, may lead to poor control at the initial stage of activating the controller due to the poor initial parameter estimates. To compensate for this effect, an extra control signal can be introduced termed "caution"(15). This is the component of the control signal that recognises that the parameter estimates may be in error. In self tuning control this function is treated in a fairly ad-hoc fashion.

If the plant can be described by a linear discrete transfer function, the coefficients of which are assumed known, then a linear feedback control can be designed that minimises some form of quadratic cost function. To avoid the complex mathematical manipulations necessary for control design based on the solution of the matrix Ricatti equation, special forms of cost function utilising only plant input/output data are minimised, for instance,

$$J1 = E\{(Y(t))^2\}$$

$$J2 = E\{(Y(t+k)-W(t))^2+\lambda.(U(t))^2\}$$

$$J3 = E\{(Y(t+k)-W(t))^2+\lambda.(U(t)-U(t-1))^2\}$$

$$J4 = E\{(P(Z^{-1}).Y(t+k)-W(t))^2+(Q'(Z^{-1}).U(t))^2\}$$

where 'k' is the integer number of sample periods that elapse before the present control signal affects the

system output, the expectation in 'J2, 3, 4' being conditional upon data acquired up to and including the present time. This expectation operation is required as the disturbance is assumed to occur randomly.

The original self tuning regulator (4) aimed to minimise the output variance 'J1', and has been used successfully for quality control in applications such as paper making and ore crushing (5). However as the control signal is not explicitly stated in the cost function, then violent control action can take place, resulting in premature wear of the final control element or wasted energy. Also, in the case of non minimum phase systems, where the plant transfer function zeroes are outside the unit disk, then instability can occur because the minimum variance controller attempts to cancel these zeroes with control law poles.

The self tuning controller due to Clarke et al (15), which is an adaption of the self tuning regulator, is based on the minimisation of cost functions 'J2,3' where 'λ' is a scalar function. This controller explicitly weights the control signal in the cost function, thus allowing output variance to be traded off against control variance, and extending the application of self tuning control to some non minimum phase systems. The convergent self tuning control law has many interpretations (38), but one that is particularly relevant is that of a control law consisting of a cascade forward path controller, and an optimal k-step-ahead

feedback predictor. For the cost function 'J2' the convergent forward path controller is simply a proportional gain, thus it is not surprising that for type 0 plant, and constant set value, then a steady error arises. The advantage of cost function 'J3' is that the convergent controller is a discrete integrator and thus the steady state error is zero. However, the presence of the integration results in a degradation of the transient performance due to the increased phase lag of the control and the necessary reduction in forward gain.

To increase the choice of desired transient performance, whilst retaining the benefit of zero steady state error, the self tuning control (17,14) that minimises cost function 'J4' was developed. The theory has been extended from the original self tuning control that minimised J2,3 to admit weighting functions '$Q'(Z^{-1})$, $P(Z^{-1})$' that are polynomials or transfer functions in the backward shift operator. This then allows convergent forward path controls to impart both phase lead and lag to the error signal, resulting in a wide range of possible transient performance.

The design of a self tuning controller that minimises this cost function is presented in the following chapters, the classical design being presented in chapters 4.1, 4.2. The application of this control presents several major difficulties, two of which are discussed and previous solutions presented.

In the past, one of the most taxing problems in applying self tuning control to practical situations, was to ensure the effective rejection of offset in the error between set value and system output. This has given rise to several different solutions, the shortcomings of two are discussed. A new form of self tuning control presented by Clarke et al (21) is described in chapter 4.3, which overcomes most of the disadvantages of previous offset rejection techniques, and is particularly simple and effective in its implementation.

4.1 Minimisation of the cost function for known plant dynamics

The concept of controller design to optimise system performance has been discussed in terms of minimisation of a given cost function. The cost function described is given by,

$$I = E\{(P(Z^{-1}).Y(t+k)-W(t))^2+(Q'(Z^{-1}).U(t))^2\} \qquad (4.1.1)$$

where the output and control weighting are transfer functions in the backward shift operator, that is,

$$P(Z^{-1}) = \frac{Pn(Z^{-1})}{Pd(Z^{-1})}$$

and

$$Q'(Z^{-1}) = \frac{Q'n(Z^{-1})}{Q'd(Z^{-1})}$$

Given that the future disturbances are unknown then it is convenient to express equation 4.1.1 in terms of

known measurements, predictions, and unknown errors. The prediction of the k-step ahead auxilliary output derived in appendix A, is given by,

$$\phi^*(t+k|t)=P(Z^{-1}).Y^*(t+k|t)=\frac{F(Z^{-1})}{C(Z^{-1}).Pn(Z^{-1})}.\phi(t)$$

$$+\frac{G'(Z^{-1})}{C(Z^{-1})}.U(t) + \delta \qquad (4.1.2)$$

and the k-step-ahead auxilliary output.

$$\phi(t+k) = \phi^*(t+k|t)+E(Z^{-1}).\xi(t+k) \qquad (4.1.3)$$

As the error term '$E(Z^{-1}).\xi(t+k)$' is uncorrelated with '$\phi(t-i)$, $W(t-i)$, $U(t-i)$' for $i \geq 0$, then equation 4.1.1 can be rewritten as,

$$I = E\{(\phi^*(t+k|t)-W(t))^2+(Q'(Z^{-1}).U(t))^2\} + \sigma^2 \qquad (4.1.4)$$

where

$$\sigma^2 = E\{(E(Z^{-1}).\xi(t+k))^2\}$$

If the expectation operator in equation 4.1.4 is conditional on data acquired up to time 't' then the expectation is deterministic, and the cost function can be minimised by minimising instead the following function with respect to the control 'U(t)'.

$$I = (\phi^*(t+k|t) -W(t))^2+(Q'(Z^{-1}).U(t))^2 + \sigma^2 \qquad (4.1.5)$$

This can be accomplished by setting the first derivative of I with respect to 'U(t)' to zero. The

first partial derivative is given by,

$$\frac{\partial I}{\partial U(t)} = 2(\phi^*(t+k|t)-W(t)).\frac{\partial \phi^*(t+k|t)}{\partial U(t)}+2.q_0.Q'(Z^{-1}).U(t) = 0 \qquad (4.1.6)$$

where 'q_0' is the first term of

$$Q'(Z^{-1}) = q_0+q_1.Z^{-1}+ \ldots.$$

and from equation 4.1.2

$$\frac{\partial \phi^*(t+k|t)}{\partial U(t)} = g_0$$

where 'g_0' is the first term of $\frac{G'(Z^{-1})}{C(Z^{-1})} = g_0+g_1.Z^{-1}+ \ldots.$

which when substituted into equation 4.1.6 gives,

$$\phi^*(t+k|t)-W(t)+\frac{q_0.Q'(Z^{-1}).U(t)}{g_0}=0 \qquad (4.1.7)$$

Defining $Q(Z^{-1})=q_0.Q'(Z^{-1})/g_0$ and rearranging equation 4.1.7 yields the control which minimises the cost function and is given by,

$$U(t) = \frac{W(t) -\phi^*(t+k|t)}{Q(Z^{-1})} \qquad (4.1.8)$$

This control is the same form as the predictive control of chapter 3. A block diagram of the system under this control is given by fig 4.1.1. The predictor in the diagram is not physically realisable, though the relationship between the variables is still valid.

From equations 4.1.8, 4.1.3, and 3.2.1 the closed loop transfer function is as follows,

$$Y(t)=\frac{B(Z^{-1}).W(t-k)+(C(Z^{-1}).Q(Z^{-1})+E(Z^{-1}).B(Z^{-1}))\xi(t)+A(Z^{-1}).Q(Z^{-1}).d}{A(Z^{-1}).Q(Z^{-1})+P(Z^{-1}).B(Z^{-1})} \qquad (4.1.9)$$

The stability, transient and steady state response of this system can be investigated by studying the roots of the characteristic equation. In many applications zero steady state error is an important feature of any control system. For constant setvalue, then for the equality,

$$E\{Y(t)\} = E\{W(t)\}$$

then it is sufficient that,

$$P(1) = 1 \text{ and } A(1).Q(1) = 0$$

These conditions may be met by the following,

$$P(Z^{-1}) = \frac{Pn(Z^{-1})}{Pd(Z^{-1})} = \frac{1}{1}$$

and

$$Q(Z^{-1}) = q_0 \frac{(1-Z^{-1})}{(1-q_1.Z^{-1})} \quad , \; q_1 \neq 1 \qquad (4.1.10)$$

From equation 4.1.8 it is apparent that the control weighting of equation 4.1.10 is equivalent to a PI form of control law in the forward loop. This form of control weighting is useful because it has intuitive appeal and through choice of q_0, q_1 allows a wide range of transient response.

If, as is likely, the value of the offset 'd' drifts (see fig 4.1.1), then if this drift is not followed by the predictor parameters, then a prediction error will occur, leading to a steady state offset. Techniques for automatic 'tracking' of drifting offsets are described in the next section.

4.2 Positional predictors

The predictor derived in appendix A is known as a positional predictor to differentiate it from an incremental predictor. An incremental predictor uses increments of input and output data to predict the future output, whereas a positional predictor uses the actual input/output data.

In chapter 4.1 it was proven that the conditional cost function can be minimised by combining a linear control law with a k-step-ahead predictor. The effectiveness of this scheme relies heavily on the accuracy of the estimates of the predictor parameters. The predictor parameters are a function of the plant parameters and the offset 'd' which are in general unknown, and may be time varying.

This chapter shows that a self tuning predictor can be devised by combining the predictive control law with a least squares parameter estimation algorithm. For the general form of cost function consisting of transfer functions then it is shown that extended least squares techniques are necessary to obtain unbiased parameters.

The major disadvantage of the self tuning positional predictor is in its inability to compensate for a time varying offset which arises from the varying mean value of the input data.

Another feature of practical data is that it is seldom contaminated with noise and disturbances that can be modelled by correlated white noise. In particular there may be intervals when a practical process is not significantly disturbed, and thus the data input to the parameter estimator may not be sufficiently "active".

The problems posed by the practical features of data obtained from a real process are discussed in this chapter, and solutions are proposed.

4.2.1 Estimation of predictor parameters

The derivation that follows will show that the predictor parameters can be estimated without bias using recursive least squares techniques, provided that the weighting functions '$P(Z^{-1})$, $Q(Z^{-1})$' are polynomials, and that the usual assumption of ideal stochastic disturbances is justified.

This is achieved by expressing the predictor in the form of the linear regression model as follows.

The optimal control law of equation 4.1.8 will be restated by defining an auxilliary function as,

$$\psi^*(t+k|t)=\phi^*(t+k|t)-W(t)+Q(Z^{-1}).U(t) \qquad (4.2.1)$$

Substitute ϕ^* from equation A.14. into equation 4.2.1 and rearrange giving the following,

$$C(Z^{-1}).\psi^{*}(t+k|t)=\frac{F(Z^{-1})}{Pn(Z^{-1})}.\phi(t)+G(Z^{-1}).U(t)-C(Z^{-1}).W(t)+\gamma \qquad (4.2.2)$$

where

$$\gamma = C(1).\delta$$

and $G(Z^{-1}) = G'(Z^{-1})+C(Z^{-1}).Q(Z^{-1})$

Let,

$$\psi(t+k)=\phi(t+k)-W(t)+Q(Z^{-1}).U(t) \qquad (4.2.3)$$

hence from equation 4.2.3 and equation 4.1.3

$$\psi(t+k)=\psi^{*}(t+k|t)+E(Z^{-1}).\xi(t+k) \qquad (4.2.4)$$

therefore substituting ψ^{*} into equation 4.2.2 and rearranging gives

$$\psi(t+k)=F(Z^{-1}).Y(t)+G(Z^{-1}).U(t)-C(Z^{-1}).W(t)$$
$$+\gamma-(C(Z^{-1})-1).\psi^{*}(t+k|t)+E(Z^{-1}).\xi(t+k) \qquad (4.2.5)$$

A delayed version of equation 4.2.5 can be expressed in the form of a linear regression model as follows,

$$\psi(t) = \underline{x}^{T}(t).\underline{\theta}+e(t) \qquad (4.2.6)$$

where

$$e(t) = E(Z^{-1}).\xi(t)-(C(Z^{-1})-1).\psi^{*}(t|t-k)$$

$$\underline{x}^{T}(t) = (Y(t-k),Y(t-k-1), \ldots ; U(t-k), U(t-k-1), \ldots .$$
$$W(t-k),W(t-k-1), \ldots .;1)$$

$$\underline{\theta}^{T} = (f_0,f_1, \ldots ; g_0,g_1, \ldots ; -C_1,-C_2, \ldots ;\gamma)$$

The offset 'γ' is accomodated by placing a "1 in the data vector" and 'γ' in the parameter vector.

As shown in chapter 2.2.4, if the error term 'e(t)' is uncorrelated with the data vector then the parameter

parameter vector can be estimated using recursive least squares without bias. As '$\psi^*(t-k|t)$' is correlated with $\underline{x}(t)$ then this is not the case and thus bias would be expected to occur. However, if the estimates of the parameters ($\hat{\theta}$) are correct, that is $\hat{\theta} = \theta$, then equation 4.1.7 and equation 4.2.1 are zero and thus the error term is given by,

$$e(t) = E(Z^{-1}).\xi(t)$$

which is uncorrelated with the data vector.

Thus if the parameter estimator is combined with a control law that ensures that 'ψ^*' is zero, that is equation 4.1.8, then it is concluded that $\hat{\theta} = \theta$ is a fixed point of the algorithm, and if this fixed point is stable then the algorithm is self-tuning (15).

The stability conditions are associated with the sensitivity of the control law at the fixed point. Thus the control must be chosen to be insensitive to small deviations in '$\hat{\theta}$', otherwise a significantly non zero 'ψ^*' will be produced giving rise to further bias in the estimates. However for this type of process control application, control signal limitations reduce the likelihood of the occurrence of sensitive control laws, and thus is an excellent candidate for self tuning control.

In an effort to increase the speed of convergence of the estimates, a forgetting factor is used to give greater

weight to more recent data. This is particularly important at start up of the controller, because the parameter estimates may be seriously in error, thus causing a very large value of ψ^* and subsequent slow convergence.

The reason for only polynomial control and output weighting functions is now obvious, because if transfer functions were allowed the noise term would be given by,

$$e(t)=Pd(Z^{-1}).Qd(Z^{-1}).(E(Z^{-1}).\xi(t)-(C(Z^{-1})-1).\psi^*(t|t-k)) \quad (4.2.2)$$

which is correlated with the data vector and thus least squares will yield biased estimates. Hence the choice of cost function is restricted by the need to obtain unbiased parameter estimates. In such instances it is more appropriate to use an extended least squares technique, as described in chapter 2.

A further disadvantage of this formulation is that the parameter vector 'θ' is a function of the user specified control weighting, '$Q(Z^{-1})$'. This function may require tuning 'on-line' which will cause the parameter estimates to vary and hence make evaluation of the variation in the cost function a difficult task.

If it is assumed that the parameters can be estimated without bias, then as the parameters tend to their true value, equation 4.2.2 tends to zero, that is the control is chosen such that,

$$\underline{x}^T(t).\hat{\underline{\theta}} = 0$$

which is satisfied for all parameter vectors of the form,

$$\hat{\theta} = \mu.\theta$$

where 'μ' is an arbitrary scalar value. In such cases the solution is said to exist on a linear manifold, and in practice the estimates will tend to wander in unison. This will not cause a problem for control signal calculation , but may cause numerical errors if the estimates become very large or very small. This can be avoided however by fixing one of the estimates at some constant value. Suitable candidates to fix are '$-\hat{C}_0$' if present, at unity or '$\hat{g}_0$' at some estimated value. The choice of '$\hat{g}_0$' is not critical, however it is restricted to be greater than 50% of its actual value 'g_0' (15), otherwise the parameter estimates may not converge.

4.2.2 Non zero mean data

In deriving the predictor equations it was assumed that the unknown disturbances to the system output are given by,

$$n(t) = \frac{C(Z^{-1})}{D(Z^{-1})}.\xi(t)+d$$

where '$\xi(t)$' is a sampled white noise sequence and 'd' is an unknown constant. This constant offset can be considered to be due to disturbances or inaccuracy in the estimation of the mean values of the input/output data. For most practical systems a linear model is only justified when considering small signal variations, and as shown in chapter 2.2, the signal variation can be

calculated by subtraction of the mean values from the input/output data. However, in practice this mean value varies due to disturbances so it is necessary to recursively estimate it or high pass filter the data to remove the low frequency components.

The parameter estimation described in chapter 4.2.1 used a "1 in the data vector" formulation of the regression model to estimate the offset, which appears at first sight to be a convenient solution. However because the data vector element is fixed at unity, then the information content of this signal is very small, and consequently the offset estimate converges very slowly relative to the other estimates. If the offset is constant then it would only have to be estimated once, and the slow convergence would not be a problem in the long term. However, in practice the offset is time varying, and in particular it may make step-like changes due to disturbances such as solar radiation or occupancy variation. The offset estimate cannot follow such rapid changes, thus leading to prediction error, and a degradation of the controlled response.

A popular alternative technique for offset rejection is to cascade the control law with a discrete integrator, and to use the error between setvalue and system output as the output data for estimation purposes. This then ensures zero mean output and input (to the integrator) providing the setvalue is constant. If however the setvalue is not constant then any variation will appear as a disturbance

to the system with two main effects:

(i) The transient response will not be the same as for a setvalue change because the variation in setvalue will appear as a disturbance.

(ii) Because the setvalue change appears as a disturbance then the parameter estimates will also be affected, the magnitude of the variation in estimates depending on the pattern of setvalue changes.

A theoretical concern is that the plant transfer function '$A(Z^{-1})$' contains a root on the unit circle due to the integrator, as does the noise transfer function '$C(Z^{-1})$'. This is reported, in principle, to cause difficulties with the convergence of the estimates (21). Fortunately this is only a problem when the plant disturbance is purely stochastic, which is not often the case in practical trials.

In Chapter 2, the offset in the data was removed by high pass filtering, which in turn led to bias due to the resultant correlated noise. The conclusion of the identification study was that given significant noise and time varying nonzero offset, then a compromise must be sought between error in the model due to offset, or that due to bias as a result of correlated noise caused by high pass filtering.

4.2.3 Data information content

In practice a forgetting factor is incorporated in any practical self tuner for two main purposes:

(i) As stated in chapter 4.2.1, the initial parameter estimates may be so seriously in error as to generate data that will cause further error in the estimates. Hence a forgetting factor is needed to give greater weight to more recent, and by assumption, more accurate data.

(ii) In practice the plant parameters vary, and the use of a forgetting factor allows slowly time varying parameters to be tracked.

It is possible to track slowly varying parameters because the forgetting factor artificially increases the magnitude of the covariance matrix at each sample. If however the information content of the input/output data is low and the setvalue is constant then the continual increase of the covariance matrix may cause the elements of the matrix to become very large in magnitude. If then, due to a setvalue change or disturbance, the information content of the data vector suddenly increases, then the large covariance matrix will cause the estimates to vary unexpectedly. This is only likely to occur after many hundred samples of data, the effect of which could conceivably lead to instability of the controlled variable.

This type of unpredictable performance must be avoided at all costs for reasons both of safety and system reliability.

These problems have been avoided in the past by the use of "jacketing software' that effectively predicts when the data information content is sufficient and/or when the parameters have changed, and activates the parameter estimator in some way at these times.

Prediction of information content may be possible if it is known when parameters are likely to change. For instance, when the setvalue changes the parameters change as a result of the new operating point and the nonlinear plant. The parameter estimator may then be given a certain number of samples to obtain new estimates before inhibiting any further changes. This type of scheme does not however directly address the problem of slowly time varying parameters.

A scheme that has some theoretical grounding (68) is the variable forgetting factor algorithm due to Fortescue et al (35). No attempt is made to derive the algorithm, a statement of the major results being sufficient to describe the operation. It is proposed that a function can be defined '$\Sigma(t)$', that describes the information content of the estimator as a function of the errors '$\varepsilon(t)$, $\varepsilon(t-1)$,....., where '$\varepsilon(t)$' is the error between the actual plant output and the predicted value based on parameter estimates, that is

$$\varepsilon(t) = Y(t)-Y^*(t|t-1)$$

Further, if '$\varepsilon(t)$' is small then this may indicate that the estimates are close to exact and/or there is little noise and disturbances acting on the system. In which case the estimates should not be varied significantly, hence the forgetting factor is increased. If '$\varepsilon(t)$' is large then this indicates that the estimates are greatly in error and/or there is significant noise and disturbances. Thus it would be wise to allow variation in the estimates and hence the forgetting factor is reduced. Based on these observations the relationship between '$\Sigma(t)$' and '$\varepsilon(t)$' can be derived as a function of the forgetting factor, and is given by,

$$\Sigma(t)=\beta(t).\Sigma(t-1)+(1-\underline{K}^T(t).\underline{x}(t)).(\varepsilon(t))^2$$

where '$\beta(t)$ is the time variable forgetting factor, '$\underline{K}(t)$' the Kalman gain vector, $\underline{x}(t)$ the data vector, and 'Σ_0' is the desired information content figure. The value of '$\beta(t)$' is altered such that the information content is held constant that is,

$$\Sigma(t) = \Sigma(t-1)= .... \Sigma_0$$

In which case the forgetting factor is given by,

$$\beta(t)=1-(1-\underline{K}^T(t).\underline{x}(t)).(\varepsilon(t))^2/ \Sigma_0 \qquad (4.2.7)$$

This calculation may generate zero or negative values hence it is usual to restrict the range of β, that is

$$\beta min < \beta < 1$$

The choice 'Σ_0' is eased if it is expressed as,

$$\Sigma_0 = \sigma m^2/(1-\beta_0)$$

where 'σm^2' is the expected measurement noise variance, and 'β_0' is the nominal forgetting factor for this noise level. This then gives some rough guidelines for the choice of 'Σ_0'. If the plant parameters are likely to vary frequently then choice of a high value for 'Σ_0' will cause the forgetting factor to reduce and hence cause a more rapid variation in estimates. For less critical applications, such as HVAC control where fast parameter tracking is not necessary or desirable then a low'Σ_0' value is to be preferred.

In the simulations and practical tests where the forgetting factor used was constant, no signs of estimator instability were revealed. This may be because the data was sufficient or the experiment was not carried out for long enough. However, several applications require that a self tuner operates for extended periods where the data may be insufficient. One such application is that of temperature or humidity control of a computer suite, in which a reliable control is of paramount importance.

4.3 The k incremental predictor

The k incremental predictor as the basis of a self tuning control law, was proposed originally by Clarke et al (21) as a means of solving the offset problem in a robust (with respect to freezing parameter estimates) manner,

whilst allowing greater flexibility in choice of control action. The effect of k-differenced data is to cause correlated residuals which may lead to bias in estimates. However, in most cases the bias caused insignificant degradation of control performance when compared to the improvement in the offset rejection properties.

This chapter presents the basic self tuning control derivation.

The k incremental predictor is formed by simply taking k differences of the positional predictor, that is, multiplying by the Z transform,

$$\Delta_k = 1-Z^{-k} \qquad (4.3.1)$$

Therefore multiplying equation A.17 by equation 4.3.1 and assuming for the moment that,

$$C(Z^{-1}) = 1$$

and for simplicity that,

$$P(Z^{-1}) = \frac{Pn(Z^{-1})}{Pd(z^{-1})} = \frac{1}{1}$$

then,

$$Y^*(t+k|t)-Y^*(t|t-k)=F(Z^{-1}).\Delta_k.Y(t)+G'(Z^{-1}).\Delta_k U(t) \qquad (4.3.2)$$

The offset is not present as its k difference is zero.

Polynomials $F(Z^{-1})$, $G'(Z^{-1})$ are unchanged by the differencing operation, as it is applied both to output and input data.

The output is a function of the prediction and the error defined as '$\hat{Y}$', such that

$$Y(t) = Y^*(t|t-k)+\hat{Y}(t|t-k)$$

Therefore substituting for Y* in equation 4.3.2 gives,

$$\Delta_k.Y(t+k)+\tilde{Y}(t|t-k)=F(Z^{-1}).\Delta_k.Y(t)+G'(Z^{-1}).\Delta_k.U(t)+\hat{Y}(t+k|t)$$

or delaying by k samples

$$\Delta_k.Y(t)+\hat{Y}(t-k|t-2k)=F(Z^{-1}).\Delta_k.Y(t-k)+G'(Z^{-1}).\Delta_k.U(t-k)+\hat{Y}(t|t-k) \qquad (4.3.3)$$

Expressing equation 4.3.3 in the form of the linear regression model

$$\phi(t) = \underline{x}^T(t).\underline{\theta}+e(t)$$

where

$$e(t) = \hat{Y}(t|t-k)$$

$$\phi(t) = \Delta_k.Y(t)+\tilde{Y}(t-k|t-2k)$$

$$\underline{x}^T(t) = (\Delta_k.Y(t-k),\Delta_k.Y(t-k-1),....;\Delta_k.U(t-k),\Delta_k.U(t-k-1),....)$$

$$\underline{\theta}^T = (f_0,f_1,....;g_0,g_1,....)$$

As the error term is uncorrelated with the data vector then the parameters can be estimated without bias by least squares methods. This is assuming that the output, given by,

$$\phi(t) = \Delta_k.Y(t) + \tilde{Y}(t-k|t-2k)$$

is accessible. The noise term is given by,

$$\tilde{Y}(t-k|t-2k) = E(Z^{-1}).\xi(t-k)$$

which is unknown as $\xi(t)$ is uncorrelated. In the spirit of extended least squares it is usual to proxy this noise term by the residual of a previous estimation. Alternatively it is possible to assert that the plant is nonstochastic and to assume the noise term is zero. If this is not justified then the error term is given by

$$e(t) = \tilde{Y}(t|t-k) - \tilde{Y}(t-k|t-2k)$$

which is correlated with the data vector and hence least squares estimates will be biased. However, if the noise level is low enough then this bias will cause insignificant degradation of controlled performance. The assertion that the plant is nonstochastic will be used in the practical experiments to be described.

As with the self tuning controls previously described, the estimated parameters are used to generate an estimate of the k step ahead output, that is,

$$\hat{Y}^*(t+k|t) = Y(t) + \hat{F}(Z^{-1}).\Delta_k.Y(t) + \hat{G}'(Z^{-1}).\Delta_k.U(t) \qquad (4.3.4)$$

and the control is calculated by,

$$U(t) = \frac{W(t) - \hat{Y}^*(t+k|t)}{Q(Z^{-1})} \qquad (4.3.5)$$

If the control weighting is the same form as equation 4.1.10, or the plant has a natural integrator, such as an incremental actuator, then steady state error will be zero for a constant setvalue and offset disturbance. Hence the form of the data vector and the control weighting ensures zero prediction error and so called 'λ' error (21).

The control weighting '$Q(Z^{-1})$' can be in transfer function format which eases the design task. Further as the predictor estimates are not a function of '$Q(Z^{-1})$', then the control weighting can be varied on line to obtain the desired response without this altering the predictor parameters.

4.4 Choice of self tuner constants and summary of algorithm

To ease understanding and for convenient reference, the variables that must be initialised in the design stage, and by the user at the commissioning stage, will now be summarised. Also the algorithm will be described in a form suitable for practical implementation using current microprocessor technology.

4.4.1 Order of estimation polynomials

The order of the estimation polynomials is primarily determined by the plant order and time delay. The order of the output weighting transfer function has some effect,

however for simplicity this function was set equal to unity. As shown in chapter 2 the second order model is a close representation of the test facility dynamics and other more realistic plant. Hence for these studies a second order plant will be assumed, unless an incremental actuator is used and then the order is increased to third. The time delay is a notional concept that comprises transport lag and the effects of several storage elements and hence varies widely from one application to another.

From equation A.8 the order of the polynomials are given by,

$$\begin{aligned} &\text{order of } \hat{F} = n-1 \\ &\text{order of } \hat{G}' = n-1+k \end{aligned} \qquad (4.4.1)$$

where 'n' is the plant order and 'k' the plant time delay in integer number of sample periods.

4.4.2 Forgetting factor and initial covariance

The nominal forgetting factor used throughout the tests is 0.998. This value is admittedly on the high side, however this is preferred so that parameter estimates do not vary unpredictably as a result of noise. Although no problems have been experienced using a fixed forgetting factor, investigations have been carried out using a variable value. In such circumstances limits are placed on the value given by the following,

$$0.125 < \beta < 1 \qquad (4.4.2)$$

and the information content measure,

$$\Sigma_0 = 0.125 \qquad (4.4.3)$$

The initial covariance is chosen as a unity magnitude diagonal matrix. This is chosen so that initial variation of parameters estimates will not be so wide as to cause unnecessary tuning-in transients. For more critical control loops where fast tuning in is required, then a large covariance may be used to cause rapid convergence of parameter estimates.

4.4.3 Initial parameter estimates

If reasonably accurate estimates are available then it would be wise to use them and start the tuner with low initial covariance, this then reflects the confidence one has in these estimates. To obtain the estimates, however, would require performing an experiment to obtain a plant model and then solving the Diophantine equation (equation A.7). This process is far too complex to be practical.

The value of $'\hat{g}_0$ is the only estimate that can be obtained easily, this being the first response to a unit step or impulse. All other parameters are set to zero. In spite of this poor initial model, reasonable results were obtained in practice.

4.4.4 Cost function weighting transfer function

The major function of the cost function weighting, besides defining the cost function, is to allow the user to select a desired transient and/or steady state response. This is governed by the roots of the characteristic equation (C.E.) given by equation 4.1.9, repeated here,

$$C.E. = A(Z^{-1}).Q(Z^{-1})+P(Z^{-1}).B(Z^{-1})=0 \qquad (4.4.4)$$

The output weighting is set to unity for simplicity, leaving only the control weighting to choose. This can be carried out initially by plotting the locus of the roots for a variation in one of the parameters of '$Q(Z^{-1})$'.

For the identified test rig model given by equation 2.3.6 that is,

$$Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})} = Z^{-k}\ \frac{(0.00071 + 0.007.Z^{-2})}{(1-1.337.Z^{-1}+0.44.Z^{-2})} \qquad (4.4.5)$$

and the control weighting given by equation 4.1.10, that is,

$$Q(Z^{-1}) = q_0.\frac{(1-Z^{-1})}{(1-q_1.Z^{-1})} \qquad (4.4.6)$$

which is the discrete equivalent of the continuous time P+I control. Two sets of PI parameters were investigated and the closed loop dominant roots were calculated. The PI parameters and dominant complex roots are given by,

for $q_0 = 0.0654,\ q_1 = 0$

$Z = 0.615 \pm j.0.245$ (4.4.7)

and $q_0 = 0.045,\ q_1 = 0.8$

$Z = 0.645 \pm j.0.267$ (4.4.8)

4.4.5 Sample time and time delay

The self tuning control law described earlier performs the functions of parameter identification and control signal calculation using the same sampled input/output data. The choice of the data sample rate is complicated because the effect on the two operations is different and often leads to conflicting requirements. Thus the choice of sample rate is a compromise between several factors.

The major influences on choice of sample rate, with regard to control performance are as follows:-

(i) plant bandwidth
(ii) bandwidth of disturbances
(iii) stroke time of actuator
(iv) computation speed

The filtering effect of the data sampler can be used to advantage by choosing a sampling frequency sufficiently high to faithfully reproduce the plant output frequencies of interest whilst attenuating the higher frequencies associated with measurement noise. The sampling theorem requires that this sample frequency be at least twice the frequency of the data of interest, however, for

practical purposes a frequency of 4 to 20 times greater than the frequencies of interest is more realistic (36). This is to reduce the delay between a setvalue change and a control signal response, and to smooth the system output response to control inputs that are applied via a zero order hold.

The significant dynamic response of the plant to the input of heat energy, is primarily determined by the thermal resistive and capacitive properties of the air in the space and of the building fabric. The frequency content of the response due to the two storage media however is very different. The bandwidth of the building fabric (walls) temperature response may be an order of magnitude smaller than the air in the space, but may contribute a significant amount to settling time for instance.

This difference of time constants is characterised in the step response by an initial steep rise of temperature, due to the thermal response of the air, followed by a less steep increase due to the thermal response of the building fabric. The significance of the two types of response is a factor of the type of application, the heating medium, and the building construction. This description of the thermal response of a building is somewhat simplistic however it does allow an understanding of the differing frequency content in the response.

To facilitate the choice of sample rate the response of the plant will be assumed to be dominated by an effective first order response. Further, the bandwidth of the response will be assumed to be approximately equal to the plant break frequency, which is calculated using the effective time constant of the first order response. Thus a practical sample rate would be between 4 and 20 times the plant break frequency that is,

$$\frac{T_C}{ST} \cong 4 \ldots 20 \qquad (4.4.9)$$

where 'T_C' is the effective time constant.

The disturbances to space temperature are due to measurement noise, unwanted inputs of heat energy, and fluctuations of primary heat source (boiler temperatures). Measurement noise coupled to detector leads is of a very high frequency relative to the plant bandwidth, thus analogue filtering at the controller inputs can be designed to reduce these effects. Disturbances due to variation in the heat energy in the plant cause temperature variations of a frequency that is restricted by the plant bandwidth. Thus the previous sample rates based on plant bandwidth are appropriate for efficient disturbance rejection. The effect of boiler temperature fluctuations on duct temperature usually requires that the space temperature controller be coupled with a duct temperature controller which can simply be a proportional analogue control. The duct control has as a setvalue the output of the space temperature controller. This scheme

reduces the fluctuations of the duct temperature at its outlet and thus makes the space temperature controller easier to tune. Solar radiation and inputs of heat due to changes in occupancy are of a varying duration. As short periods of increased temperature may not cause discomfort, then in an effort to reduce the wear on valves and actuators, it is wise to increase the sample length to that just tolerable for adequate disturbance rejection.

The stroke time of an actuator is the time taken for the actuator to alter its output position by 100%. For positional control the output of the controller is proportional to the steady state actuator position. If the sample time is smaller than the stroke time of the actuator then the actuator position may not be able to follow the control signal if large changes occur. This is particularly a problem if both heating and cooling stages are being controlled because a situation may arise where both heating and cooling are applied simultaneously. This situation can also arise with incremental control. With this form of control, the actuator position is changed by an amount proportional to the control signal, thus if the sample time is too small then part of the control signal is lost. This loss of control signal, known as windup, can lead to a degradation of the response, and several techniques have been devised (7) to compensate for such effects. The most straight forward of the techniques is to make the sample time equal to stroke time.

The computational requirements to calculate the control signal for an implicit self tuner are relatively small in terms of time and data storage. This is because the implicit self tuner identifies the predictor parameters directly, thus avoiding the solution of the Diophantine equation (equation A.7), and making the control signal calculation a trivial matter relative to the parameter estimation calculations.

The major influence on choice of sample time, with regard to parameter estimation efficiency are as follows:-

(i) Parameter accuracy
(ii) Numerical problems
(iii) Computation speed

If a linear continuous system is sampled at a fixed rate, then a Z transform transfer function can be derived to describe the evolution of the samples, which will be exact if the input to the linear system is Z transformable. It is well known that as the sample rate is increased the roots of the system Z transfer function migrate to the unity point in the Z plane, and eventually cluster around it. Thus for a high enough frequency the difference in the roots will be numerically small, and only very small changes in the roots are necessary to cause significant change in the transient response of the model. This increased model performance sensitivity to parameter error requires that sample time must not be chosen too small, otherwise estimation accuracy is critical. If however the

sample time is too large then not enough information of the correct frequency content will be present in the samples to adequately describe the dynamics of the plant. Thus the sample time must not be chosen too large or too small or the resulting model will be inaccurate.

If the sample time is chosen too small then numerical problems can arise in the parameter estimation algorithm (53), because the difference equations describing the system output became approximately linear dependent. Thus the sample time must be chosen large enough to avoid such problems.

The parameter identification described in chapter 2 used the same data, and hence sample time, as estimated to be adequate for impulse response determination via crosscorrelation. The degree of correlation between model response for the identified response and crosscorrelation illustrate the adequacy of the model and thus the choice of sample time. The sample time was chosen for crosscorrelation so that a 63 stage PRBS would take approximately the plant settling time to be completed. This length PRBS is commonly used on chemical processes to derive simplified models for control (91) design and is likely to be adequate for HVAC control considering the similarity of plant. The settling time is approximately equal to 5 effective time constants, hence the sample time is given by,

$$\frac{ST}{Tc} = \frac{5}{63} \cong \frac{1}{12} \qquad (4.4.10)$$

As this sample rate is also adequate in terms of the controller requirements previously described it will be used in the following experiments.

The major computational burden of the self tuning algorithm is the least squares parameter estimator, and as this is performed every sample, then this operation is the significant factor in determining the maximum sample rate. Using a low cost microprocessor based self tuning controller it has been reported (15) that it is possible to perform the calculations to update 10 parameter estimates in less than one second. Considering that the time constants of applications requiring self tuning control are likely to be measured in hours, then the computation time is unlikely to be a restricting consideration in the choice of sample time.

4.4.6 Summary of self-tuning control algorithm

The k incremental self tuning control algorithm will now be summarised in a systematic format suitable for programming on a digital computer.

1. Initialise variables and input user specified constants

2. Read system output

3. Form data vector $\underline{x}(t)$

$$\underline{x}^T(t) = (\Delta_k.Y(t-k), \Delta_k.Y(t-k-1), \ldots \Delta_k.U(t-k), \Delta_k.U(t-k-1), \ldots)$$

$$\Delta_k.Y(t-k) = Y(t-k) - Y(t-2.k)$$

$$\Delta_k.U(t-k) = U(t-k) - U(t-2.k)$$

4. Replace $Y(t)$ by $\Delta_k.Y(t)$ in the recursive least squares algorithm given by equation 2.2.16. Solve equation 2.2.16 to obtain the regression coefficient vector given by,

$$\underline{\hat{\theta}} = (\hat{f}_0, \hat{f}_1, \ldots \hat{f}_{n-1}; g_0, g_1, \ldots g_{n-1-k})$$

5. Generate the control signal and hence the implicit prediction of the k step ahead output. That is

$$U(t) = \frac{W(t) - \hat{Y}^*(t+k|t)}{Q(Z^{-1})}$$

and as

$$Y^*(t+k|t) = Y(t) + \hat{F}(Z^{-1}).\Delta_k.Y(t) + \hat{G}'.\Delta_k.U(t)$$

then,

$$U(t) = \frac{(W(t) - Y(t) - \hat{F}(Z^{-1}).\Delta_k.Y(t) - (\hat{G}'(Z^{-1}) - \hat{g}_0).\Delta_k.U(t))}{Q(Z^{-1}) + \hat{g}_0}$$

6. Output control signal and wait for sample period to elapse.

7. Goto 1.

This summarises the basic algorithm, however in practice, this algorithm would exist as one module in a more

extensive system of procedures. The other functions would include:

1. Forgetting factor calculation.

2. Imposition of control signal limits and compensation for external limitations of control action.

3. Limit control. It is common practice to specify high and low limits on controlled variables such as duct temp, and if the variable exceeds the limits then some form of limit control scheme is activated.

4.5 Simulated examples

To illustrate the performance of the algorithm previously described, a series of simulated experiments will be carried out. These experiments seek to show that the assumptions that were made to ease the design procedure were fully justified.

The plant to be simulated is the general linear model given by,

$$Y(t) = Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})}.U(t)+\frac{C(Z^{-1}).\xi(t)}{A(Z^{-1})}+d \qquad (4.5.1)$$

where the symbols are explained in chapter 2.2.4, and the values are given by,

$$\frac{B(Z^{-1})}{A(Z^{-1})} = \frac{(0.00071+0.007.Z^{-1})}{(1-1.337.Z^{-1}+0.44.Z^{-2})} \qquad (4.5.2)$$

$$C(Z^{-1}) = 1 \qquad (4.5.3)$$

k is given various integer values > 0

d is constant or a ramp function

4.5.1 Stochastic measurements

The basic assumption of the derivation of chapter 4.3 is that the plant is non-stochastic. This assumption although possibly untrue is justified if the effects of making such an assumption is not to cause unacceptable degradation of the closed loop system response.

The experimental conditions are:

$E\{(\xi(t))^2\} = 0.01$

$Q(Z^{-1}) = 0.0654$

$W(t)$ = square wave

$d = 10$

The two following cases will be considered:

1. output for estimator = $Y(t)-Y(t-k)$
2. output for estimator = $Y(t)-Y(t-k)+\hat{Y}(t-k|t-2k)$

   which if k = 1 then

   $\hat{Y}(t-k|t-2.k) = \xi(t-1)$

Recall equation 4.3.3, hence for case 1 the estimated parameters will be biased, whereas for case 2 the parameters can be estimated without bias.

The exact predictor parameters can be calculated from equation A.7 and are given by

$$F(Z^{-1}) = 1.337-0.44.Z^{-1}$$

$$G'(Z^{-1}) = 0.00071+0.007.Z^{-1}$$

The estimated 'F' parameters and the system output response for cases 1 and 2 are shown in figs 4.5.1 and 4.5.2 respectively. The bias for case 2 in fig 4.5.1 is quite obvious whereas the output responses for cases 1 and 2 'appear' similar. This apparent insensitivity is due partly to the low time delay and also to the nature of the control law.

This experiment shows that given exact knowledge of the system noise, unbiased estimated may be obtained using recursive least squares techniques. Further, if, as is likely, the noise is unknown then under certain conditions the response will be unaffected by the biased estimates.

For the rest of the experiments the estimator system output will be given by case 1, that is

$$\text{output} = Y(t)-Y(t-k)$$

4.5.2 Time varying offset

The previous experiment was carried out under proportional control and hence a steady state error would be expected to occur. If the control is proportional plus integral, then for zero steady state error the prediction error must also be zero. As shown by chapter 4.3 the advantage of incremental predictors is the robust offset elimination. Robustness in this case is the ability to

have offset elimination despite biased predictor parameters.

The plant given by equation 4.5.2 combined with a proportional plus integral control is equivalent to a unity feedback type 1 system. For such a system the steady state error is zero for a constant setvalue and offset, and is constant for a ramped offset and constant setvalue.

This experiment illustrates that despite biased predictor parameters a constant mean error is achieved with a ramped offset.

The control weighting is given by,

$$Q(Z^{-1}) = 0.045.\frac{(1-Z^{-1})}{(1-0.8.Z^{-1})} \qquad (4.5.4)$$

and the output response is given by fig 4.5.3 for the experiment conditions:

d is ramped as illustrated

k = 1

$E\{(\xi(t))\}^2 = 0$

The slight error is constant, after the initial transient, and changes sign at the same time as the rate of change of offset. Hence for constant offset the error would be zero.

For the same conditions as for previous experiment except that,

$E\{(\xi(t))^2\} = 0.01$

the response of fig 4.5.4 was obtained. The response shows that the mean of the error does not increase with offset variation, thus for constant offset and setvalue the mean error would be zero.

The simulated plant input response shows that due to the power limits the transient output response will not be as given by the characteristic equation. This is shown in figs 4.5.3 and 4.5.4 by a reduced rate of output response for reducing setvalue.

4.5.3 Misassignment of time delay

Up till this point it has been implicitly assumed that the value of the time delay is known exactly. In practice only a rough estimate will be known, hence it is important to know what the effects of misassignment of this important factor are. To make comparison between over and under-estimation possible the time delay is set to 5. The same experimental conditions are used as for chapter 4.5.2 except for the following,

fig 4.5.5a estimated delay = 5

fig 4.5.5b estimated delay = 4

fig 4.5.5c estimated delay = 6

k = 5

the order of $\hat{G}'$ is increased correspondingly

Comparing the second positive going transients, it can

be proposed that the effect of underestimation is to cause a more oscillatory response, overestimation causing a more sluggish response. In which case overestimation is to be preferred in practice.

The initial large overshoot is due to a combination of poor initial predictor parameter estimates and the plant time delay. The control signal takes 'k' samples before its effect is realised at the plant output, hence before the estimator can make a reliable variation in the predictor parameters, 'k' samples must elapse. During this time the control signal is generated using a poor predictor thus causing the large initial overshoot. To avoid such problems a 'CAUTIOUS' control may be proposed. This would take the form of an increased control weighting for the first 2.k samples. In practice this may not be necessary due to the fact that power limitations will restrict the initial transient rate of response.

Thus in practice some leeway is allowed in the estimation of time delay, however, if the delay is sufficiently underestimated then a large transient may be generated or an unstable response will result.

4.5.4 Nonlinear plant effects

In general, the small-signal gain of valves and heat exchangers is operating point dependent. Therefore, if the operating point changes then so do the plant model

parameters. As operating points may vary as a result of setvalue change or disturbances, then it is important to investigate how closely the estimator parameters follow the model parameters.

To simulate the effects of an operating point dependent nonlinearity, the input to the plant simulation "U´(t)" is given as a 2nd order nonlinear **function** of the control output "U(t)", as follows,

$$U'(t) = 0.007813.U^2(t)+9.25 \qquad (4.5.5)$$

see fig 4.5.6

The experimental conditions are as follows:

k = 5, estimated delay = 5

d = 10

$E\{(\xi(t))^2\} = (0.01)^2$

$Q(Z^{-1})$ given by equation 4.5.4

$\beta = 0.998$, $P(0) = I.1$

The response is given by fig 4.5.7. The initial transient is characterised by a larger overshoot due mainly to the poor predictor parameter estimates. The further transients show that for increased setvalue, and hence operating point, the response is more oscillatory than for a reduced setvalue. This is a direct result of the increased small signal gain at higher operating points. The ideal closed loop response is not achieved because during such a transient the plant parameters vary too rapidly for the estimator to follow. However, the estimates can be made to vary more rapidly by artificially increasing the

magnitude of the covariance matrix. Alternatively, the covariance matrix can be reset to a large magnitude diagonal matrix when a setvalue change takes place. The algorithm, due to Fortescue (35) described in chapter 4.2, varies the forgetting factor to keep the information content of the self tuner constant. That is, if the data is unexciting then the effective memory length tends to infinity, and if the data is exciting then the effective memory length reduces. An occasion when data is exciting is when a setvalue change occurs, in which case if the plant parameters change due to operating point, then the reduced forgetting factor and hence increased covariance will cause a fast alteration in the parameter estimates.

The Fortescue algorithm was used with the same experimental conditions as for fig 4.5.7 and,

$\Sigma_0 = 0.125$

$0.1 \leqslant \beta \leqslant 1$

and the response is given by fig 4.5.8a,b.

For comparison the same experiment was repeated with the same experimental conditions as for fig 4.5.7 except

$\beta = 1$

P = 1*I reset every setvalue change

The response is given by fig 4.5.9.

These results show that attempting to adapt to fast changes in plant parameters is unlikely to improve the

response. This is due mainly to significant delay in control signals and uncertainty in measurements due to noise and disturbances. The variable forgetting factor algorithm will continue to be used, but its parameters will be altered to ensure that the estimates will only vary slowly in response to plant parameter changes, that is,

$$\Sigma_0 = 0.125$$

$$0.9 \leqslant \beta \leqslant 1$$

5 Comparison of conventional and self tuning control of the heating system

To assist in the evaluation of the self tuning control, the design, commissioning and long term operation of a conventional controller will be described in parallel with the self tuning control. The form of conventional control chosen is the two term proportional plus integral, or PI, the reasons for this choice having been explained in chapter 3.

The performance will be compared for the following:

(i) Ease of commissioning

(ii) Sensitivity of the response to plant parameter changes

These aspects will be shown to be closely linked.

The commissioning of a controller is defined here as the adjustment of the control variables by a commissioning engineer, on the site where the plant is operating. This is necessary because the initial "factory" settings of the control variables may not be suitable for all plant applications. Thus, for reasons of economy, the commissioning procedures should not be excessively time consuming, or require special equipment or highly skilled personnel to carry it out.

The procedures to be presented are not proposed to be the best possible ways of commissioning controls for this application, but two of many possible techniques.

The linear model derived in chapter 2 is only valid about a particular operating point. Hence if this point changes due to load variations or setvalue changes then the plant dynamics will also change. The sensitivity of the response to the variation in the plant dynamics will determine how often the control has to be recommissioned and/or how rapidly the quality of performance is degraded.

Due to the subjective nature of the realisation of comfort conditions, the making of a definitive judgement on control performance quality is a particularly difficult task. However, it is without dispute that an attractive quality of a control scheme is that its use results in a nett reduction of fuel useage. The results of a recent survey (72) of domestic central heating systems, revealed

a link between thermostat differential and net fuel useage. It was shown that for minimum thermostat differential and hence minimum cyclic variation in temperature, the householders set the thermostat at a lower value to give comfort, than did the householders who had a larger thermostat differential. This then resulted in a lower mean temperature and hence lower energy consumption. Apparently this indicates that the thermostat was set in relation to the minimum temperature realised rather than maximum or mean.

Although thermostat differential does not apply to the modulating controls under investigation, cyclic variation can result for particular plant or control conditions. Thus a basis for judging control performance may be that the response is nonoscillatory when subjected to a load or setvalue change. This quality however is exhibited by a zero gain control, amongst others. Hence it is usual to also specify the maximum time to steady state. This will depend on the power limitations and also the magnitude of the setvalue or disturbance change. A further consideration is that highly active control signals cause undue wear of valves and motor equipment.

Thus due to the complex nature of the problem it is possible only to make fairly informal judgements on response quality.

The experiments will be carried out using simulated and actual data from the test rig, the simulated data being generated from the identified model given by equation 2.3.6. This model indicates that there is insignificant delay to the control output, this being due to the small transport lag of the heating medium. In practice the transport lag will be greater due to longer ducts, storage effects of ancillary equipment, and perhaps most significant of all, the mixing processes within the air space. This delay may be further increased depending on the positioning of the detector within the air space. Thus to simulate the practical plant more accurately a time delay will be inserted between calculation of control signal and output to the rig.

The time delay is chosen as 160 seconds, which relative to the dominant time constant of approximately 400 seconds, may be considered large by some standards and small by others.

For these studies the actuator valve and heat exchanger will be assumed to constitute an operating point dependent gain.

5.1 Commissioning of controllers

The commissioning procedures to be presented are based on the information obtained from the process reaction curve. This is the response of the plant to a step change in

control output of a sufficient magnitude to overcome the noise and disturbances on the plant. This experiment is chosen because, beside taking less time than most other techniques, it is simple to carry out and requires no special equipment.

The plant characteristics obtained from the process reaction curve are as follows:

(i) steady state gain
(ii) effective dead time
(iii) dominant time constant, assuming a first order response
(iv) storage time

The steady state gain taken from the process reaction curve is a good estimate of the small signal gain if the small signal gain is approximately constant throughout the range of the control signal variation.

The combined effects of time delays and short time constant storage effects is assumed to be adequately modelled by the effective dead time.

If the plant is first order dominated then the dominant time constant is given by the time taken by the output to reach 63% of its final value minus the effective dead time.

The storage time is defined as the area between the reaction curve and the final steady state output divided by the magnitude of the steady state change of output which for a first order response is equal to the time constant plus time delay.

The process reaction curve obtained for a 100% step variation in control signal is drawn by fig 5.1.1, and the plant characteristics are given by,

$$\begin{aligned} K &= 0.079 \quad \text{steady state gain} \\ Tps &= 700 \text{ secs} \quad \text{storage time} \\ Td &= 266 \text{ secs} \quad \text{effective dead time} \\ Tc &= 440 \text{ secs} \quad \text{time constant} \end{aligned} \tag{5.1.1}$$

5.1.1 Conventional Control

A convenient method of tuning controls is from a compiled chart of control parameters and plant characteristics that satisfy a given performance criterion. The plant characteristics can be those obtained from the process reaction curve as previously described.

The tuning charts due to Fertik (33) will be used in these experiments. These charts were compiled for a simulated plant consisting of two lags plus time delay. Tuning of the PI control algorithm was for minimum ITAE and overshoot to a step set value change constrained to less than 8% of set-value change. The values of control parameters obtained from such charts are unlikely to give

an optimum response, thus further manual tuning is required. This may require experienced personnel and a great deal of time and hence add considerably to the installation and commissioning costs.

Further tuning will be carried out by analysing the closed loop response to a setvalue step increase of $4^{o}C$. This will result in an approximate mean control output of 50% The control will be tuned to give a transient that is minimally oscillatory and fast, given the power limitations. A variation in setvalue is used to tune the control, as causing a representative disturbance to the space temperature is a particularly difficult task.

The plant characteristics of equation 5.1.1 are transferred to the tuning charts giving the PI control in the Laplace operator as follows,

$$C(S) = 9.3.(1+\frac{0.002}{S}) = \frac{U(S)}{C(S)} \qquad (5.1.2)$$

As the control is carried out by digital computer, this equation is more directly useful in Z transform format. The Z transform of equation 5.1.2 does not exist because the error 'e(t)' is a continuous signal, hence the mapping from 'S' to 'Z' planes can only be carried out approximately. One of the many techniques for approximate mapping is the Tustin Bilinear transformation. This is simply carried out by substituting 'S' in equation 5.1.2 by the following,

$$S = \frac{2}{ST} \cdot \frac{(1-Z^{-1})}{(1+Z^{-1})} \qquad (5.1.3)$$

The transformation effectively maps the entire 'jw' axis in the 'S' plane into the unit circle in the Z plane. This mapping, not surprisingly, can cause frequency distortion if the frequency of sampling is high enough. However, for frequencies as low as one tenth the break frequency no appreciable distortion should occur. That is, the sample time is given by,

$$ST \cong \frac{\text{time constant}}{10}$$

therefore from equation 5.1.1 a sample time of 40 seconds would be adequate.

Making the substitution for 'S' in equation 5.1.2 by equation 5.1.3 with ST = 40, the Z transform control is given by,

$$C(Z^{-1}) = 9.672.\frac{(1-0.923.Z^{-1})}{(1-Z^{-1})} = \frac{U(S)}{C(S)} \qquad (5.1.4)$$

where $e(S) = W(S) - Y(S)$

'W(S)' is the setvalue and 'Y(s)' the plant output. The response for a $4^{o}C$ increase in setvalue is given by fig 5.1.2. The response is characterised by an initial reduction of the error due to proportional control action, and then a slow reduction of the error to zero due to integral control action, and hence is a particularly sluggish response.

In an effort to increase the speed of the response, the gain is increased by 50%, such that the control law in terms of 'S' is,

$$C(S) = 14.(1+\frac{0.002}{S}) \qquad (5.1.5)$$

and by the Tustin transform,

$$C(Z^{-1}) = 14.6.\frac{(1-0.923.Z^{-1})}{(1-Z^{-1})} \qquad (5.1.6)$$

The closed loop step response for a control given by equation 5.1.6 is given by fig 5.1.3.

In an effort to reduce the oscillation the gain of the control is reduced, but to increase the speed at which the error is reduced then the integral gain is increased.

For a control given by,

$$C(S) = 12.7.(1+\frac{0.0025}{S}) \qquad (5.1.7)$$

and the digital equivalent,

$$C(Z^{-1}) = 13.34.\frac{(1-0.905.Z^{-1})}{(1-Z^{-1})} \qquad (5.1.8)$$

the closed loop step response is given by fig 5.1.4.

Increasing the integral term and reducing the gain once more gives the control,

$$C(S) = 11.(1+\frac{0.003}{S}) \qquad (5.1.9)$$

and the digital equivalent,

$$C(Z^{-1}) = 11.66.\frac{(1-0.89.Z^{-1})}{(1-Z^{-1})} \qquad (5.1.10)$$

the closed loop step response is given by fig 5.1.5. This response will be assumed to be the best that can be achieved, in terms of speed and oscillation, for the limited time and resources available to a commissioning engineer. The initial undershoot in fig 5.1.5 is a remnant of a previous test, this is proven by there being no similar undershoot illustrated in the control signal response of fig 5.1.6.

If the plant closely approximates a first order lag plus time delay then there is a particularly simple tuning technique due to Haalman (46) that can be used for commissioning. For a plant with Laplace transfer function,

$$G(S) = Kg.\frac{\exp(-S.Td)}{(1+S.Tc)} = \frac{Y(S)}{U(S)} \qquad (5.1.11)$$

the control is given by,

$$C(S) = Kc.(1+\frac{1}{Ti.S})$$

where

$$Kc = \frac{2.Tc}{Kg.3.Td}$$

and $Ti = Tc$

If the process reaction curve is used to obtain the plant characteristics then from equation 5.1.1, the control is as follows,

$$C(S) = 13.26.(1+\frac{0.0023}{S}) \qquad (5.1.12)$$

Compared to the final tuned control equation 5.1.9, this appears to be a reasonable estimate. This technique is only appropriate if the time delay is significant relative to the time constant, otherwise very large control gains can be calculated. However, this technique illustrates a useful principle that is relevant for many applications.

5.1.2 Self tuning control

The self tuning controller derived in chapter 4 has the structure of a linear forward path controller operating on the error between setvalue and a prediction of the k-step-ahead output. Which, if the predictor parameter estimates are unbiased, gives a closed loop characteristic equation as follows,

$$C.E. = A(Z^{-1}).Q(Z^{-1})+B(Z^{-1}) = 0 \qquad (5.1.13)$$

Hence the closed loop response is fixed by a suitable choice of the forward path controller '$1/Q(Z^{-1})$'. This choice is greatly eased by the removal from the C.E. of the time delay, which ideally should make the commissioning procedure less time consuming due to the reduced sensitivity of the response.

The plant will be assumed to be a first order lag plus time delay and the commissioning procedure based on the technique due to Haalman (46). The first order approximate plant model given by equation 5.1.11 has a Z transform because the input (U(t)' is sampled, which for a 40 second sample time is given by,

$$G(Z^{-1}) = \frac{0.0072.Z^{-k}}{(1-Z^{-1}.0.91)} = Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})} \qquad (5.1.14)$$

For a PI type control given by,

$$\frac{1}{Q(Z^{-1})} = \frac{1}{q_0}.\frac{(1-q_1.Z^{-1})}{(1-Z^{-1})} \qquad (5.1.15)$$

then following the Haalman principle of cancelling the plant pole,

$$q_1 = 0.91 \qquad (5.1.16)$$

The closed loop characteristic equation is then obtained from equations 5.1.13, 5.1.14, 5.1.15 and is given by,

$$C.E. = (1-0.91.Z^{-1}).q_0.\frac{(1-Z^{-1})}{(1-q_1.Z^{-1})} + 0.0072 = 0$$

which upon substitution of equation 5.1.16

$$C.E. = (\frac{(q_0+0.0072)}{q_0} - Z^{-1}).\frac{1}{q_0} \qquad (5.1.17)$$

Thus the closed loop transient response, which is dictated by the characteristic equation, will be first order dominated with time constant adjustable via 'q_0'. Following the Haalman principle further would mean setting 'q_0' to zero as the predictor has reduced the delay to zero. This would not be realistic, instead a value of

'q_0' will be chosen that is typical for HVAC applications. The control gain is usually denoted as the proportional band, which is defined as the number of degree centigrade error to cause the control variable to alter by its full range. For HVAC control the range of proportional band is generally from $2^{o}C$ to $10^{o}C$.

The value of 'q_0' will be arbitrarily chosen to give a proportional band of $5^{o}C$. Thus for proportional only control, 'q_0' is given by,

$$q_0 = \frac{5^{o}C}{(120^{o}-10^{o})} = 0.045^{o}C/^{o} \text{ firing angle}$$

The control is then as follows,

$$\frac{1}{Q(Z^{-1})} = 22.2.\frac{(1-0.91.Z^{-1})}{(1-Z^{-1})} \qquad (5.1.18)$$

Comparing this control law to the conventional control law given by equation 5.1.10 it is obvious that the presence of the predictor has allowed an increased in forward gain of approximately 100%. However, it is yet to be seen how this control law performs in practice.

The self tuner requires several constants to be specified before the control can be activated, these include control weighting, sample time, time delay, number of predictor parameters, and initial parameter estimates. These are fully described and summarised in chapter 4.4.4.

The control weighting has been derived based on a sample time of 40 seconds, this being chosen because of the

satisfactory identification results of chapter 2, and control results of chapter 5.1.1. This value will be used throughout these experiments.

An important factor is the estimated time delay, as was shown in chapter 4.5. The time delay in this case will be estimated by the number of samples elapsed before the first non zero change in the process reaction curve occurs. From fig 5.1.1 the number of samples is 7, equivalent to a delay of 280 seconds.

The number of predictor parameters estimated is not critical, as will be shown, and hence the maximum number that this RLS algorithm can handle, that is ten, was chosen. To be precise, 2F and 8G parameter were estimated.

As pointed out in chapter 4.4.3, initial estimates for the predictor parameters are not easily obtained, hence the estimates are set to zero except for 'g_0' which is obtained from fig 5.1.1 as,

$$g_0 = 0.0031$$

This initial predictor estimate will be so poor that the initial transient is likely to be significantly degraded. Thus to illustrate the 'tuned' setvalue change response, a second setvalue change will be made 50 samples after the first. The first setvalue change is $+4^{o}C$, as for the

conventional control, and the second + 2°C.

All other parameters are as described in chapter 4.4.

The setvalue change response is as shown by fig 5.1.7 and the control signal response by fig 5.1.8. As previously experienced, the initial transient is poor but the error is rapidly reduced once the parameter estimator has had sufficient data. The second transient is a close approximation to an overdamped response, showing insignificant oscillation, and the rate of error reduction compares well to that obtained using conventional control fig 5.1.5.

Thus it appears from this experiment that this particularly simple commissioning procedure gives a control performance that compares well to conventional control without the need for further manual tuning. Further tests will be carried out in chapter 5.2 to investigate the effect on the setvalue response of changes in the control weighting gains, and to further confirm the results of this chapter.

5.1.3 Commissioning schemes for self tuning controls

The commissioning scheme considered in chapter 5.1.2, required information that was obtained from the process reaction curve. In practice the commissioning engineer would neither have the time nor the equipment to carry out such a test, so a simpler commissioning procedure is

required before the self tuning controller can be considered viable. Several schemes are now discussed which require a varying amount of intervention by the commissioning engineer.

If the control design procedure described in chapter 5.1.2 is pursued, then the control parameters that must be specified can be estimated from the process reaction curve. The data obtained is the effective time delay and the storage time of the plant. As the process reaction curve cannot be measured manually then some other means of estimating this data must be devised. The accurate estimation of time delay has been shown to be important for establishing a stable closed loop response, and as it is difficult to analytically calculate this parameter, then some form of experiment will be needed to aid its estimation. Fortunately, the time delay is easily measured, and it is relatively constant therefore it need only be measured infrequently, for instance, once every 24 hours at plant start up. The time delay is constant because it depends largely on the ventilation rate which is usually fixed. It can simply be measured as the time taken for the system output to reach some predefined level from the time of start up of the plant, as shown by fig 5.1.1. This threshold value can then define the initial value of the parameter estimate '$\hat{g}_0$'. For the time between start up and time delay calculation a "cautious" form of control law could be used, such as

a proportional control with a fairly large proportional band. Furthermore, this initial control could be used for a fixed time after the time delay while parameter estimation takes place, thus reducing the possibility of poor predictor estimates causing a large initial temperature overshoot and a resultant waste of energy.

The estimation of the plant storage time from the process reaction curve would require a very large experimentation time and thus is not a practical proposition. Furthermore, a scheme that relied on the measurement of the storage time from the process reaction curve would have two fundamental disadvantages:

(i) Requires the cooperation of the customer to start up the plant, and the experiment will require the use of fuel whilst the building is unoccupied which can be considered as wasted.

(ii) If the primary energy supply system (boiler) has not been allowed to reach steady state, then the temperature response will be influenced by the boiler response and hence an inaccurate result may be obtained.

Thus considering the practical difficulty of performing the experiment and the likely inaccuracy of the results, it may be more appropriate to estimate the storage time from an analytical approach such as the calculation of the "fill time" of the space. The accuracy of this

approach was shown in chapter 2.2.2 to be rather poor, however, if the predictor is accurate then it will reduce the sensitivity of the system response to such poor estimates. The fill time is however likely to be a sufficiently accurate indication of plant storage time, such that an initial estimate of sample time can be estimated from the approximate relationship between storage time and sample time given by equation 4.4.10. This is because a wide range of sample time will be adequate for both control and identification, see equation 4.4.9. Thus the initial stages of such a commissioning scheme are listed as follows:-

(i) The commissioning engineer estimates the plant fill time and inputs this value, plus a chosen value of proportional band. The proportional band may be nominally $5^{o}C$ but can be set in the range $2^{o}C$ to $10^{o}C$ depending on the desired speed of response.

(ii) The controller then sets a sample time based on the plant fill time and controls the plant to achieve the desired setvalue. This initial "cautious" control can be simply proportional control with a wide proportional band which will ensure a stable nonoscillatory response.

(iii) When the system output is detected to have started to vary in response to the initial control signals, then the time delay is calculated as the elapsed number of samples

after the initial control signal was applied. The predictor structure can then be defined based on this delay and an assumed adequate order of plant model.

(iv) The cautious control may then continue to be used for a given number of samples, such that the initially poor prediction parameters can be improved by the parameter estimator before the self tuning control is activated.

The advantage of this scheme is its computational simplicity and the intuitive appeal of the foreground control commissioning procedure. The operation of the background self tuning predictor is, in principle, to allow control design and controller performance to continue as if the time delay is zero. The disadvantages are that the plant is required to be modelled as a first order time delay system, and that the commissioning engineer must make judgements that may require his experience of plant response.

A self tuning control that requires minimal commissioning, has been proposed by Dexter et al (29) for application to radiator heating systems. The controller only requires manual estimation of sample time, which is related to an estimate of the plant storage time as previously described. Estimation of the weighting function "$Q(Z^{-1})$' is carried out by the controller using a "hill climbing" technique, such that the weighting function is altered until a given

cost function is minimised. The cost function is the Integral of Absolute Error (IAE) evaluated over a 24 hour period, and depending on its value relative to the value for the previous 24 hour period, a decision is made whether to increase or decrease one of the weighting function parameters. Estimation of time delay is achieved in a similar way as that previously described. The advantage of this scheme is that the minimum of commissioning has to be carried out. However as the weighting function is only altered once every 24 hours then the rate of convergence of the control law is likely to be slow. Also, in the event of unsatisfactory performance of the controller, the only manually alterable parameter is the sample time, and it is not obvious how this should be varied to improve the performance.

The self commissioning, self tuning control is completely autonomous, the only way it can be overridden is by disconnecting the control signal. Thus in the event of unsatisfactory control performance, the customer can only disconnect the controller and contact the manufacturer. This may result in inconvenience, wasted energy due to unsatisfactory control response, and as the site visit of the commissioning engineer is likely to be charged to the manufacturer under a warranty agreement, then a subsequent increase in commissioning cost. To avoid such problems it is usual to allow the customer to have some measure of influence over the control parameters.

This is especially the case in process control applications where the disruption caused by a controller failure is likely to be more serious in terms of safety and cost than a similar failure would be in a HVAC application. These considerations have motivated manufacturers of adaptive process controls to produce controllers that calculate the optimum (in some sense) control parameters, and then merely "advise" the customer of these new parameters by displaying them. The customer may then instruct the controller to use these values if it is thought that an improved performance may result. Alternatively, new control parameters may be input to the controller manually. In such a situation "hill climbing" or deterministic search procedures cannot be used to alter the control parameters, as they require feedback of the controlled response. Instead, it is more likely that the controller will identify a linear model of the plant, and from these model parameters the control parameters can be calculated. Such a commissioning scheme will now be presented.

The commissiong scheme described in chapter 5.1.2, which is based on the design principle due to Haalman (46), can be further simplified if the requirement for input of forward path controller gain and integral time was removed. This may be achieved if the plant gain and storage time were estimated by the controller on-line, by assuming a first order plant and using least squares identification techniques as described in chapter 2. The

plant may be assumed to be adequately described by the first order system given by,

$$\Delta_k Y(t) = \hat{K}.\Delta_k U(t-k) + \hat{a}.\Delta_k Y(t) \qquad (5.1.19)$$

where Y(t) is the system output, U(t) the control signal for positional control, and $\hat{K}$, $\hat{a}$ the unknown parameters. The data in equation 5.1.19 is also used for predictor parameter estimation so this does not need to be calculated, furthermore, the least squares parameter estimator may be 'shared' between predictor estimation and first order plant estimation. Once the parameters $\hat{K}$, $\hat{a}$ have been estimated, the "advised" forward path controller parameters can then be calculated and displayed. The forward gain can be calculated to give a desired loop gain 'Kd', where from equation 5.1.15 the control gain '$1/q_0$' is given by,

$$\frac{1}{q_0} = \frac{Kd}{\hat{K}}$$

and following the Haalman design procedure then from equation 5.1.15 'ql' is given by,

$$ql = \hat{a}$$

The customer can then decide whether to continue with the previous control parameters, use the advised values, or manually input some new values.

The role of the background self tuning predictor is to reduce the sensitivity of the system response to

variation in the plant or control parameters, thus the adjustment of the foreground conventional control is eased as inappropriate control parameter settings are tolerated more readily.

5.2 Sensitivity of controlled response to forward gain changes

This chapter illustrates the likely effects on the commissioned response of the variation of forward gain, which in general is operating point dependent. The forward gain is chosen in particular because this factor is likely to vary most significantly and have the most detrimental effect on the controlled response.

The operating point, and hence the gain, may vary either as a result of a setvalue change, or external influences such as ambient temperature . Such that a control may be commissioned to have a desired transient response for a given setvalue change and ambient temperature, however for a different ambient temperature or setvalue change the transient response may not be as desired. This phenomena is shown in chapter 5.2.1 to seriously affect the conventional control commissioned response, however the self tuning control response is largely unaffected. Changes in forward gain due to the control gain is also shown not to cause significant variation in the self tuning control performance. However it is shown to be a convenient means of reducing the variance of the control signal, and thus reduce the likely wear on valves and actuators.

The role of the temperature controller in a practical HVAC system is mainly to regulate the temperature conditions despite the varying heat load on the system. Simulating practical disturbances is very difficult thus controls are usually commissioned to give a desired transient response to a setvalue change, and the response to a disturbance is inferred from this transient.

The transient response of the system to a disturbance is thus also affected by operating point dependent gain. These affects are described in chapter 5.2.2 for disturbances comprising of correlated white noise and step functions. The self tuning control is shown to have a largely constant disturbance response despite a slowly time varying forward gain, the conventional control response is shown to vary, and for significant variation in gain become unstable. These results being obtained from simulation and the test facility.

Reduction of the number of predictor parameters does not seriously affect the performance, however significant underestimation of time delay is shown to reduce the stability of the control.

5.2.1 Transient response to setvalue variation

In practice the HVAC control system will have a supervisory timing function which will define a setvalue for the period for which the building is occupied, and for the nonoccupancy period which is usually to provide

background heating to prevent condensation, and hence the control performance is not critical during this time. For experimental purposes the occupancy and nonoccupancy periods will equal 100 samples, and the occupancy and nonoccupancy setvalues will be 6°C and 2°C above ambient temperature respectively. The test is carried out over a long time period, that is 1000 samples, so that the fully tuned self tuning control performance can be investigated.

The experiments are all performed on the test facility using the same experimental conditions as for the commissioning tests of chapter 5.1. The same conventional control parameters, are used, the self tuning control is altered so that parameter estimation is inhibited in the nonoccupancy period. As described in chapter 4.5, stopping parameter estimation helps to avoid unnecessary variation in predictor parameters when transferring from nonoccupancy to occupancy and hence degradation of transient response.

The only disturbances to the response are caused by ambient temperature variation and measurement noise caused by electrical switching. This measurement noise occurs only during occupancy which conveniently simulates small disturbances, it is not large enough however to significantly alter the operating point.

The conventional control response is given by fig 5.2.1 and heat exchanger power by equation 5.1.10. The temperature response is characterised by an overshoot of approximately 1°C for positive changes in setvalue and a very slow error reduction for negative changes in setvalue. Although the overshoot is much greater than for the initially commissioned response (fig 5.1.5), the peak variation in the heat exchanger power input is approximately the same for figs 5.1.6 and 5.2.2. Further, the steady state variation in power input is approximately 40% greater for fig 5.1.6 than fig 5.2.2. Thus the initial overshoot and disparity in steady state power variation may be attributed to the variation in steady state gain of the heat exchanger with respect to operating point. Referring to the duct temperature to heat exchanger power input characteristic of fig 2.2.1, and assuming that this constitutes the major nonlinearity in the relationship between exhaust temperature and heat exchanger power input. Then the steady state increase in duct temperature for the commissioned response, fig 5.1.6, will be approximately 9°C for an increase of heat exchanger power of 40%. However, for a similar increase in duct and exhaust temperature, fig 5.2.2 shows that the steady state variation in heat exchanger power is only 25% for the increase in setvalue from 2°C to 6°C above ambient. Therefore, as suggested the initially commissioned performance is not repeated due to the variation in operating point and hence forward gain.

The self tuning control response is given by fig 5.2.3 for the control weighting given by,

$$\frac{1}{Q(Z^{-1})} = 22.22.\frac{(1-0.91.Z^{-1})}{(1-Z^{-1})} \qquad (5.2.1)$$

The initial transient response is not as well damped as the second or third, this being due to the initially poor predictor parameter estimates. However the initial transient has less overshoot than for the conventional control response, despite the same experimental conditions and the short time for the predictor parameter to "tune in". Subsequent transient response for positive setvalue change are well damped as initially commissioned, the increased plant gain does not cause an increased overshoot. The transient responses for negative setvalue change are much slower than for positive setvalue change, this being due to the reduced plant gain at low heat exchanger input values, as shown by fig 2.2.1, and the inhibited self tuning for nonoccupancy periods. During the transient between nonoccupancy and occupancy periods the parameter estimator is active, and during this time the plant gain varies significantly from one operating point to another and this occurs very rapidly. Although the estimated parameters can only follow slowly time varying parameters, this initial change in forward gain may cause the estimates to vary slightly resulting in a degraded transient performance. However, any differences in the positive transient between fig 5.2.3 and the commissioned response fig 5.1.7 are insignificant.

So that the sensitivity of the control response can be investigated with respect to choice of weighting function, the two following weighting functions were used with exactly the same experimental conditions as for the previous self tuning control response.

$$\frac{1}{Q(Z^{-1})} = 27.78.\frac{(1-0.91.Z^{-1})}{(1-Z^{-1})} \qquad (5.2.2)$$

$$\frac{1}{Q(Z^{-1})} = 16.67.\frac{(1-0.91.Z^{-1})}{(1-Z^{-1})} \qquad (5.2.3)$$

Output and heat exchanger responses are given by figs 5.2.5, 5.2.6 and figs 5.2.7, 5.2.8 for control weighting given by equation 5.2.2 and equation 5.2.3 respectively.

For this significant variation in control gain, equivalent to 50% of the initially tuned gain, the output response remains well damped and the speed of error reduction compares well with that given by conventional control. The effect of the variation in control gain can most easily be seen in the heat exchanger power input response. For the higher gain control, fig 5.2.6, the power variation is noticeably greater than for the low gain control, fig 5.2.8. Thus the simple commissioning procedure is effective, in that the response is approximately first order dominated with speed of response, or control signal variance, dictated by the control weighting gain.

5.2.2 Disturbance rejection

The significant disturbances that a HVAC control system must compensate for are heat loads due to occupants, machinery and solar gain, also cooling loads due to wind effects, and humidity loads due mainly to occupants. Disturbances that pose a problem to controls are those due to occupancy and direct solar exchange through windows, which cause an instantaneous addition of heat energy to the space and a subsequently rapid variation in the temperature. The solar radiation and wind incident on the building opaque structure cause a low frequency disturbance to space temperature, such that the total effects of the disturbances appear as a noise or pulse like function superimposed on a time varying, non-zero mean value.

Although the high frequency disturbances may only cause a small, approximately linear, variation in operating point, the low frequency mean value will cause the mean operating point to vary significantly throughout the heating season. Thus the sensitivity of response to disturbances for variation in operating point is of much interest.

As a preliminary to practical experiment, simulated results will be obtained for a plant model given by,

$$Z^{-k}\frac{B(Z^{-1})}{A(Z^{-1})} = Z^{-5}\frac{(0.00071+0.007.Z^{-1})}{(1-1.337.Z^{-1}+0.44.Z^{-2})} \qquad (5.2.4)$$

as previously described in chapter 4.5.

To simulate an operating point dependent nonlinearity the control output to the plant is defined as a nonlinear function of the calculated control output as shown by fig 4.5.6.

The simulated disturbance, which will be added to the simulated plant output, will comprise a zero mean noise-like or pulse-like function added to a low frequency ramp. The ramp function will be arranged to cause the mean control signal to vary through its full range. The responses for the two types of control are compared in terms of tendency to oscillate after a disturbance.

To illustrate the stabilising effect of the self tuning control, a disturbance will be used of the form,

$$x(t) = \frac{C(Z^{-1})}{A(Z^{-1})}.\xi(t)+d(t) \qquad (5.2.5)$$

where '$\xi(t)$' is a member of an uncorrelated noise sequence, $E\{(\xi(t))^2\} = 0.01$

$$C(Z^{-1}) = A(Z^{-1})$$

and 'd(t)' is a ramp function shown in the figures to follow.

The conventional control law and the self tuning control weighting are as derived in chapter 5.1 for the test facility. These controls should adequately perform with the plant model of equation 5.2.4 as this model was identified from the test facility using techniques described in chapter 2.

For the second order nonlinear forward gain, the conventional and self tuning control responses are given by figs 5.2.9 and 5.2.10. Due to the nature of the response it is particularly difficult to determine whether the response is continuously oscillating. The responses for the third order nonlinear gain are given by figs 5.2.11 and 5.2.12. The variation of the output for self tuning control does not apparently increase for reduced disturbance mean and hence increased gain. However for conventional control the response eventually becomes unstable as the forward gain increases. Thus these results show the stabilising effect of the self tuning control law.

However, such an exciting disturbance is uncommon for this particular application, and the nonlinearity is very severe. Such conditions may only occur in a few applications and hence it is relevant to consider more likely situations.

The input of heat energy to an air space from solar radiation and/or increased occupancy, may be adequately simulated by a step change in the output of the plant. How quickly the disturbance effects are reduced determines how much energy is unnecessarily used and/or how comfortable the occupants feel. The response of the two control laws to a constant mean stepped disturbance is shown by fig 5.2.13 and 5.2.14. Besides the stepped disturbance a white noise term is also added to the

output, of variance $(0.01)^2$, and the forward gain is linear. The response to the disturbances for both types of control is nonoscillatory and error is reduced to zero rapidly, as would be expected from the linear plant.

So that the effects of operating point dependent gain may be investigated, the forward gain will be given by the second order nonlinearity, and the disturbance mean value will have a negative ramp such that the mean operating point increases over a period of 1000 samples. The results are given by figs 5.2.15 and 5.2.16. The conventional control response to the disturbance step function is shown to increase in oscillation as the mean disturbance reduces, this trend continuing until the oscillations persist for the 100 samples. However, the response for self tuning control is largely unchanged throughout the 1000 samples, thus illustrating the insensitivity of the self tuning controlled response to slowly time varying perameter changes.

In practice solar radiation passing through windows etc will transmit heat to the air mainly by heating incident surfaces and these in turn heating the air flowing over them. The effects of solar radiation can be simulated in the test rig by placing an incandescent lamp within the air space. The lamp will heat the air by incident radiation on surfaces and also by air flowing directly over the lamp surface. The temperature response of the exhaust air to switching a 60 Watt light, on and off every 100 samples is shown by fig 5.2.17. To obtain the

same experimental conditions in practice, as were used for the previous simulated experiments, it would be necessary to control the ambient temperature so as to cause a variation of the operating point as required. As this facility is not possible with this experimental equipment, then to simulate the effect of variation of operating point dependent gain, an approximately constant operating point will be used with an artificially varied control signal gain.

This gain simply multiplies the calculated control signal by a factor that varies linearly from 0.5 to 4 over a period of 1024 samples. The results of regulating the temperature in the test rig for the described experimental conditions for conventional and self tuning control is shown by figs 5.2.18 and 5.2.19 respectively. These results are very similar to the simulated results given by fig 5.2.15 and 5.2.16, in that the self tuning control is largely insensitive to slow variations in forward gain, whereas the conventional control response becomes oscillatory for increased gain.

It is relevant at this stage to investigate the response for self tuning control to less than ideal experimental conditions. In particular, the effect of greatly reduced number of predictor parameters, and the misassignment of plant time delay.

Theoretically, for a first order plant and system input delayed by 7 samples, then from equation 4.4.1, the number of predictor parameters is 9, 1F and 8G parameters. Up till present 2F and 8G parameters have been estimated and excellent results have been obtained. Reducing the number of estimated parameters will increase the speed of execution of the algorithm and reduce the necessary data storage requirements, which if the algorithm is to be implemented on a single chip microcontroller (28,29), are important factors. Alternatively, as any estimate of the order of a practical system is an underestimate, then it is relevant to investigate the likely effects of this underestimation on the controlled performance.

The previous experiment for self-tuning control which gave the disturbance response of fig 5.2.19 is repeated with the same experimental conditions except for the number of predictor parameters estimated. The number of parameters is reduced from 10 to 4, that is, 2F and 2G parameters, and the response is given by fig 5.2.20. Compared to fig 5.2.19, this result illustrates similar insensitivity to gain changes and hence shows that the order of the predictor may be liberally chosen and hence greatly eases commissioning and implementation.

As previously stated, the choice of estimated time delay is important, as an underestimation may lead to instability. Thus, as the time delay may be miscalculated at commissioning or may change due to an unexpected change

in the ventilation rate, it is necessary to investigate the sensitivity of the response to misassignment.

The previous experiment which gave the disturbance response of fig 5.2.19 is repeated with the same experimental conditions except for estimated time delay and number of predictor parameters. The time delay will be estimated as 4 samples, as opposed to 7 samples calculated from the step response. Further, due to the underestimation, the number of 'G' parameters is reduced by 3 to 5. The response is given by fig 5.2.21 and is shown to be more oscillatory then fig 5.2.19. However, despite the gross underestimation of time delay the response is still stable and compares very favourably with the conventional response, fig 5.2.18, for the same conditions. Overestimation of time delay is to be preferred in practice, as it is unlikely to cause a more oscillatory response, as shown in chapter 4.5.3.

5.3 Conclusion of comparison

The results of this comparison of conventional (PI) and self tuning (k incremental predictor) control will now be summarised.

The performance of the controls and the design procedures were compared on the basis of:

(i) Ease of commissioning

(ii) Control of nonlinear plant

The former basis of comparison is problematical because a different technique for commissioning was used for the 2 types of control. Hence it may be argued that the conventional control commissioning procedure was not the 'best' that could be devised. However, the success of the self tuning control to give a closed loop response close to that predicted, without the need for manual tuning shows great advantage when compared with the conventional control commissioning results.

The problem of comparison of commissioning procedures may to a certain extent be avoided by comparing instead the sensitivity of the commissioned responses to variation in the plant and controller parameters. As the commissioning procedure can only practically be based on setvalue change response, then a measure of the 'commissionability' is the sensitivity of the transient response to setvalues of varying magnitude. The forward gain of this and most other plant is operating point dependent, therefore the plant forward gain will vary with setvalue. The response of the conventional control to a setvalue other than that designed for is given by figs 5.2.1 and 5.2.2. The response is characterised by large overshoots due to increased forward gain. The variation in forward gain due to operating point change is very rapid, and from chapter 4.5.4 it would be expected that the response of the self tuner would be

seriously degraded. However, the self tuning control response is shown, figs 5.2.3 and 5.2.4, to be similar to that designed for, in that minimal overshoot occurs and the response is fast and unoscillatory. This apparent insensitivity is further illustrated by figs 5.2.5 and 5.2.6 and figs 5.2.7 and 5.2.8, these being the response for the control weighting gain altered by 50% of the commissioned gain value. The temperature responses are very similar to fig 5.2.3 despite the gain change, however, the control signal responses of figs 5.2.6 and 5.2.8 show that for reduced gain the control signal activity is noticeably reduced.

Up till the present time the major role of the HVAC control system has been regulatory, hence transient response to setvalue change has been rather unimportant, due partly to the occurrence of setvalue changes during nonoccupancy periods. The comparison of the regulatory qualities of the control techniques is based on the sensitivity of the disturbance response to changes in the plant gain. The plant gain changes as a result of the time varying non zero mean disturbance and the fact that the plant gain is operating point dependent. Initially, simulated results were obtained, and the first of these considered the stability of the response when the disturbance comprises a low frequency ramp function plus a white noise signal. The results, figs 5.2.11 and 5.2.12 were obtained for a rather severe 3rd order nonlinearity and it was shown that for such a

nonlinearity the conventional control response can become unstable whereas the self tuning control response remained stable and substantially unde graded.

In practice disturbances are less exciting than a white noise signal and may be simulated by deterministic functions such as a step function. The control of simulated plant for a stepped disturbance is compared for the 2 types of control, the comparison being based on how quickly the disturbance effects are reduced and how oscillatory is the prevailing transient. The results for linear plant and constant mean disturbance are given by figs 5.2.13 and 5.2.14, and are similar for both types of control. This experiment is repeated but for a second order nonlinear plant gain and a ramped mean disturbance, and the results are given by figs 5.2.15 and 5.2.16. These results show that the conventional control response degrades to an unacceptable degree due to the gain variation, whereas the self tuning control response appears unchanged.

Practical validation of these results was obtained using the test rig, with a 60 Watt incandescent lamp in the air space providing the simulated solar disturbance. The response for the uncontrolled exhaust temperature is shown by fig 5.2.17. The varying forward gain being provided by multiplying the control signal by a factor that varies linearly from ½ to 4 in 1024 samples, and the

results are given by fig 5.2.18 and 5.2.19.

These results illustrate the ability of the self tuning control to compensate for slowly time varying plant parameters, in particular plant forward gain. In the case of rapidly time varying parameters, such as that due to operating point changes, the self tuning control response may be degraded, but from the experiments the response is still closer to the commissioned response than is the conventional control. Thus there appears to be significant advantage in using the self tuning control described, when plant parameters vary and the plant comprises significant time delay.

An important feature of any control system is that it is tolerant of inaccurate initial settings, especially with respect to stability of response. For instance, if a stable response can be obtained for initially inaccura te control parameters, then on-line adjustments can be made to the parameters while the control loop is actually in use. The 2 major user definable settings of a predictor are the number of coefficients used to make the prediction, and the estimated time delay. The results of fig 5.2.20 and 5.2.21 show that these features may be significantly inaccurate and the self tuning predictor will still operate satisfactorily.

6. Effect of severe plant nonlinearities

The basis of the modelling techniques described in chapter 2 is that the plant can be described by a linear model if the variation in the variables is sufficiently small. The assumption is justified for the particular plant conditions investigated, but may not be for more practically realisable experimental conditions.

For most practical plant there are component parts which are difficult to linearly model regardless of the magnitude of the associated variables. In HVAC systems such nonlinearities are mainly due to the valve and actuator combination, and are characterised by:

(i) Saturation

(ii) Hysterisis or backlash

These characteristics are most commonly encountered with the incremental type actuator and hence this type will be investigated in the experiments to follow.

The actuator, which may be pneumatically or electro-mechanically powered receives a signal from the controller which indicates that the actuator should open the valve or close it at a constant rate, the amount of time this signal is present represents the magnitude of the control signal. For simulation purposes the actuator will be considered to have as input, the control signal, and as output the sum of the control signal. The accumulated control signal is limited to the maximum and

minimum firing angle values of 120 and 10, the maximum control signal magnitude being limited to the sample time. Thus the actuator input and output are given by $U'(t)$ and $U(t)$ respectively where,

$$U(t) = U'(t) + U(t-1)$$

$$10 \leqslant U(t) \leqslant 120$$

$$-ST \leqslant U'(t) \leqslant ST$$

If control signals do not saturate then the Z transfer function of the actuator and hence the plant has a pole at unity, in which case the controller does not require such a pole to assure zero steady state error. The inclusion of the actuator pole in the plant transfer function increases the plant order by one, and hence increases the predictor parameter number by two, as in chapter 4.4.1.

The plant dynamics are simulated by the identified plant model given by equation 5.2.4. The disturbances to the plant will be simulated by equation 5.2.5 where

$$C(Z^{-1}) = A(Z^{-1})$$

$$E\{(\xi(t))^2\} = (0.01)^2$$

and $d(t)$ is either a constant or low frequency ramp. As in chapter 5 the control law used with simulated data is that obtained by the commissioning procedure of chapter 5.1, except in this case the control law pole is supplied by the actuator in the plant transfer function. Thus the conventional control law is given by,

$$C(Z^{-1}) = 11.66.(1-0.89.Z^{-1})$$

and the control weighting,

$$1/Q(Z^{-1}) = 22.2.(1.0.91.Z^{-1})$$

and the simulated plant transfer function,

$$Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})} = \frac{Z^{-5}.(0.0007+0.0071.Z^{-1})}{(1-Z^{-1}).(1-1.33.Z^{-1}+0.44.Z^{-2})}$$

Due to the increased plant order the number of estimated predicted parameters is increased to 3F and 7G parameters.

6.1 Control signal and actuator saturation

The control signal magnitude is equal to the length of time for which the actuator moves the valve stem. Thus the maximum control signal magnitude is equal to the sample time. For large setvalue or load change, the calculated control signal magnitude may be larger than the sample time, thus the response of the system will not be the same as for the linear system in which signal saturation does not occur. This degradation in response due to loss of control signal is known as "wind-up". Several compensation techniques for such practical limitations are possible (7), the simplest of which is to save the excess control signal greater than the sample time, and use it next sample. This type of scheme is used for the conventional and self tuning control experiments to follow.

Although "wind-up" can degrade the response, it can be compensated for, and more importantly, it effects both

self tuning and conventional control. This is because the parameter estimator is supplied with valid information even when control signal saturation occurs, thus the saturation effects only the control calculation and not the identification.

The advantage of the incremental actuator with respect to the positional actuator is primarily cost. The electromechanical positional actuator is effectively an incremental actuator with added electronic circuitry that senses an error between actuator position and control signal, and drives the actuator motor to reduce this error. In normal practice the control signal output range is matched to the actuator input range, such that a 100% variation in control signal causes a 100% variation in actuator position. Thus for positional actuation the position of the actuator is known precisely at all times.

The incremental actuator on the other hand has no feedback of position and hence it is not possible to predict the position to any degree of accuracy. In particular it is unknown whether the actuator position is at either extremes of its range, and hence whether further control signals will have any effect on the actuator position. The control signal information used by the parameter estimator may then be in error if the actuator is saturated.

The loss of control signal information due to practical limitations such as saturation has been recognised (94) as one of the major obstacles to the successful application of self tuning control. In a study of such practical problems, Zanker and Wellstead (94) suggest that "jacketing software" must be used to ensure that the self tuner receives only valid data. Such a technique is proposed in a report by Clarke (15), in which the parameters are only estimated if the actuator is known to be unsaturated. Detection of actuator saturation is carried out by testing the estimate '$\hat{g}_0$', which is proportional to forward gain and stopping parameter estimation if the estimate is less than some chosen minimum value. However, before considering such a scheme it must be established that saturation is likely to occur, and that the effects are such as to warrant compensation.

The results to follow illustrate where saturation occurs and its likely effects on the self tuning control performance.

Saturation commonly occurs in heating system when the demand for heating or cooling is greater than the plant can supply. This can occur when the ambient air temperature falls below the minimum designed value and the plant is unable to make up the necessary extra heat. To simulate this effect the disturbance 'd(t)' of equation 5.2.4 will be varied at a constant rate such as to cause saturation for at least one simulated

occupancy period, that is, 250 samples.

Two distinct situations will be investigated as follows:

(i) Saturation occurring after the initial occupancy period when the estimates have largely converged.

(ii) Saturation occurring during the initial transient when large parameter variations are possible, and it is expected, when the most degradation to the subsequent transients will take place due to the inaccurate estimates.

The response of the system will be judged by inspection of the system output when the actuator has just become unsaturated. This is the period when the parameter estimates obtained during saturation will effect the response to the most significant degree.

The response to a stepped setvalue and ramped offset, such that saturation occurs after the initial tuning-in transient, is given by figs 6.1.1, 6.1.2, 6.1.3. As can be seen from fig 6.12 the maximum simulated power input has been reduced from 100% to 45% to cause valve saturation at the output temperatures of interest. This response was taken with the unusual condition that the saturation was signalled to the estimator so that only control signals that caused actuator movement were used.

This response can then be used as a reference to compare with further responses. The technique of altering the control signal to reflect the fact that the actuator is saturated is known as "reflecting the limit".

The response for the limit not reflected is given by figs 6.1.4, 6.1.5, 6.1.6. These results show the effects of saturation occurring for one complete occupancy period, the estimator being halted for nonoccupancy periods. The transient response at 750 samples is the first after saturation, and as can be seen by comparing fig 6.1.1 and fig 6.1.4 the system response has not been significantly degraded. This is mainly due to the slow variation of estimated parameters while the actuator was saturated. The rate at which parameter estimates vary is governed primarily by the forgetting factor, the closer to unity the forgetting factor is the slower the estimate variation.

Thus it appears that to avoid the detrimental effects on estimates due to control signal loss, it may be sufficient to merely choose a value of forgetting factor that allows an insignificant variation in estimates for the likely time that the saturation exists. If however it becomes apparent that the parameter estimates are restricted to an unacceptable degree due to the high forgetting factor, then it may be necessary to consider some other scheme to avoid the effects of saturation, such as estimation of whether actuator is saturated or not.

The system output responses for saturation occurring during the initial transient are shown by fig 6.1.7 for limit reflected, and by fig 6.1.8 for limit not reflected. As previously illustrated, the saturation appears not to have a significantly deterimental effect on the system response, despite the saturation occurring during the initial transient.

A further reason for the insignificant degradation of the response is that although the control signals do not affect the system output during saturation, the output is altered by disturbances and measurement noise. Thus the parameter estimator does not identify a plant that has effectively a zero gain value. Alternatively if the noise and disturbances are insignificant then the information content of the output signal will be low, and thus the forgetting factor will be set to approximately unity. This then will inhibit the variation of the estimates and hence preserve the performance of the controller.

6.2 Actuator and valve hysteresis

The nonlinearities associated with the actuator and valve linkage are simplified and shown in block diagram form by figs 6.2.1, 6.2.2, 6.2.3. The dead zone nonlinearity for an electromechanical actuator is due to the electromagnetic hysterisis of the motor and the stiction of the gearbox and linkage, such that if a

control signal is not of sufficient duration, then the motor will not have developed enough torque to move the actuator. This will result in a steady state error unless effective compensation is introduced. It is common practice to compensate for deadzone by accumulating small control signals until the sum exceeds the estimated deadzone and then to output the sum to the actuator. This scheme combined with the anti-windup scheme presented in chapter 6.1 is described by the Psuedo code listing given by fig 6.2.4 and is used in the experiments to follow with a deadzone value of 2 seconds. The deadzone will be assumed to be known exactly hence its effect will be similar for conventional and for self tuning control. This assumption is justified because in practice the deadzone is small, and its effects are insignificant when compared to the destabilising effects of hysterisis.

For electromechanical actuation the hysterisis is caused by gearmeshing and play in the valve linkage. This is not such a problem for the positional actuator because the fedback position signal is derived from a potentiometer attached to the final drive shaft, hence gearmeshing effects are removed by the feedback loop. The linkage between final drive shaft and valve stem may introduce hysterisis but it will not be as great as for the incremental actuator.

The actuators used in HVAC systems exhibit hysterisis, while in the normal operating range, typically of 1% of

maximum actuator position (78), however, due to wear this value may rise. A worst case value of hysterisis of 10% will be assumed for this type of plant and will be used in the experiments to follow.

The results are obtained by simulating the plant by equation 5.2.4, and using a linear relationship between control signal and input to plant. As hysterisis will degrade the performance of both conventional and self tuning control, then the conventional control response is included for comparison. For a constant mean disturbance the response for self tuning and conventional control is given by figs 6.2.5 and 6.2.6 respectively.

The response for self tuning control is very oscillatory but improves, that is, becomes less oscillatory, as time increases. The conventional control on the other hand is relatively sluggish, this difference being primarily due to the higher gain of the self tuning control. Therefore, the response can be made less oscillatory by reducing the forward gain by increasing the control weighting.

Alternatively, if the value of the hysterisis is known then its degrading effects can be reduced by adding a compensating factor to the control signal whenever it changes sign. To be exact, for this experiment the factor equals ± 20%. To illustrate this scheme the same

experiment as previous was repeated, except that the control signal was altered to compensate for the hysterisis, and the results are given by fig 6.2.7 and 6.2.8. The reduction in the oscillation for fig 6.2.8 illustrates the effectiveness of this scheme.

Thus for perfect knowledge of the husterisis its detrimental effects can be completely removed. In practice, only as estimate of the hysterisis will be known thus it is important to see the effect of over and underestimation of the hysterisis.

To illustrate the effects of inaccurate estimation of hysterisis upon the self tuning control the previous experiment will be repeated with the estimated hysterisis equal to 7.5% and 12.5%, an over and under-estimation of 2.5%. The results, fig 6.2.9 and fig 6.2.10, are not significantly different from the case for perfect estimation, fig 6.2.8, however the response for overestimation is slightly less oscillatory than for underestimation. Hence for these experimental conditions, estimation of hysterisis to within ± 2.5% can be assumed to cause insignificant degradation to the controller performance.

If a hysterisis value of 1% is assumed to be typical for HVAC systems, then it appears likely that self tuning control can be applied to temperature control without the need for compensation for hysterisis effects.

7. Conclusions

The application of self tuning control to warm air heating systems has been investigated. The work is sufficiently generalised such that the results are relevant to other forms of temperature and humidity control.

This research was motivated by a desire to make the commissioning of temperature controls less time consuming, improve the commissioned control performance, and retain this performance despite changes in the thermal characteristics of the heating plant.

The use of self tuning control to achieve these objectives is made economically viable by the use of the ever increasing computational power and storage capacity of microprocessor devices. Self tuning control has advantages over most other forms of adaptive and optimal control, in that the storage and computational requirements can, in certain situations, be similar to more conventional control techniques, such as the ubiquitous PID algorithm. Thus in principle the same range of facilities can be offered by a self tuning controller, as a conventional controller of a similar cost, thus making self tuning control an attractive proposition.

The original theory of self tuning control is well developed (15,5,17,14,89), current research exists to

solve problems that are particular to certain applications (21), and to make self tuning control useful to a wider range of applications. Research (28) has shown that a simplified form of self tuning control algorithm can be implemented using a single chip microcontroller, thus extending the range of likely applications to the less capital intensive schemes such as HVAC control.

The role of this research is to investigate the application of self tuning control by addressing the particular problems that arise due to the nature of HVAC plant. As it is proposed that self tuning control replace the existing conventional control techniques, then the performance of the different forms of control are compared where applicable.

For the majority of HVAC systems the physical state of the air in the space is altered by the introduction and mixing of a volume of air of a different physical state. Thus in determining the state of the air space in response to various stimuli, the air mixing process is a particularly important factor. However, the analytical treatment of the air mixing process is intractable for an arbitrary air space. Thus to allow the air space dynamics to be modelled and investigated in general, a small scale test facility was constructed which exhibits some of the physical properties of a full scale occupied air space.

An important property that the results from the test facility has demonstrated, is the spatial distribution of the air space dynamics. This gives rise to different temperature responses at different points in the space, and in turn demonstrates the importance of positioning the temperature detector at a point where it will indicate the temperature most representative of that realised by the occupants of the space.

The heat input to the test facility is controlled by electronic means, so to simulate the characteristics of a real plant the nonlinearities of the valves and actuators are simulated by the control computer. This allowed investigation of the effect of these characteristics, and revealed that they have a significant influence on the overall system response.

Although the test facility may be considered a theoretically ideal air space, with low thermal mass apart from that due to air, high level of insulation, and large volumetric flow rate of supply air, its dynamic temperature response does not compare well with a simplified model frequently used to describe air space temperature dynamics, as shown in chapter 2.2.2. This indicates the inaccurate assumptions made to derive the model, and further justifies the use of an experimental test facility to describe the dynamics of a real plant.

The aim of the modelling study is to present techniques to derive low order models of heating systems using results obtained from the test facility. These low order models are then used to design conventional controls, and used as the basis for self tuning control.

The low order model is obtained by analysis and simplification of the governing physical equations of the heating system such that the structure of the model is derived, the unknown coefficients of the model are then obtained from experimental data using several techniques. These include fitting a first order model to the step response, using crosscorrelation to obtain the weighting sequence, and least squares techniques to obtain a transfer function representation of the system. Estimation of the weighting sequence by crosscorrelation of measured output and the PRBS input has the advantages that little a-priori knowledge of the plant and noise dynamics is required and accurate results are obtained despite the presence of correlated disturbances to the system output. The weighting sequence estimates, which are optimal in the least error variance sense, are then used to indicate the accuracy of the identified transfer functions by comparing with their unit pulse responses.

Errors in the weighting sequence estimates, due to drift of mean temperature values and asymmetrical nonlinearities, are shown to be easily reduced to an insignificant level. The drift in mean temperature is due to ambient

temperature variation, thus its effect can be removed by subtaction of the ambient variation from the measured output. The appearance of secondary peaks in the weighting equence is not typical of linear systems, the presence of these peaks indicates asymmetrical nonlinearities such as differing rates of heating and coooling. The use of an Inverse Repeat PRBS cancels the second order nonlinear effects, which is shown to be sufficient to completely remove the secondary peaks. Although the weighting sequence estimates are optimal, this type of model is not convenient for control design.

The design of controller is most easily facilitated by parametric models such as the discrete transfer function, the parameters of which are identified using least squares techniques. The recursive least squares algorithm is well developed, and is suitable for implementation using a short word length number representation as commonly used in microprocessor based controllers. A floating point number representation is used in this study, however successful results have been obtained using fixed point representations (28), thus further simplication of the algorithm is possible.

Measurement noise and ambient temperature variation cause a distrbance to the measured air space temperature, such that models identified using recursive least squares techniques are biased, and the unit pulse responses do not compare well with the weighting sequence. To reduce bias, generalised least squares and extended

least square techniques are used and result in a more accurate representation. These techniques rely on approximating the noise correlation transfer function by an autoregression, the order of which may be very high if the transfer function has zeroes on the unit circle. This occurs if the plant contains a natural integrator or the data vector is differenced to remove constant mean values. An alternative to differencing data is to identify a constant, however this leads to very slow convergence of estimates, especially if the mean value varies.

These results have shown that a second order model of the test facility can be identified, which compares well with the weighting sequence obtained by crosscorrelation if the order of the noise autoregression is sufficiently high.

The accuracy of a model used for controller design is a function of how the predicted control performance compares to actual performance. The results of closed loop control tests showed that the models identified using least squares techniques could be used to predict the plant input and output variance, and the results compared well with that measured from experimental data. However, the first order model obtained from the process reaction curve was used to predict the variances and the results were only slightly less accurate than for the least squares models despite the low correlation between the unit pulse response of this model, and the weighting

sequence. Thus for these experimental conditions, if the control performance is specified in terms of input and output variance, then a biased model such as the first order model may be perfectly adequate for controller design.

In modelling HVAC systems, a pure time delay is used to model the combined effects of the delay due to the transport lag of the heating medium, air mixing phenomena, and heat storage effects of ancilliary equipments. The test facility model exhibits negligible pure time delay due to the efficient mixing of air in the space, and the high volumetric air flow rate. Thus to stimulate the effects that cause time delay, the control signal was delayed by an integer number of sample periods before being output to the heater control circuit.

The destabilising effects of time delay are shown in chapter 3 for a proportional unity feedback control, and a first order lag plus time delay plant. The results show that for a given system response, the sensitivity of this response to the controller or plant gain increases as the ratio of time delay to time lag increases. Thus as the time delay increases, the adjustment of the control gain to ensure a particular system response becomes more difficult, and the designed performance is more easily affected by changes in the plant gain.

The "lumping" of storage and transport lag effects into a pure time delay is a convenience for modelling and, in principle, leads to a particularly simple control design technique. This control comprises of two distinct parts:

(i) An exact output predictor in the feedback path.

(ii) A forward path control designed as if the time delay is zero.

If the prediction horizon is exactly equal to the plant time delay, then the delay is removed from the characteristic equation, and part (ii) will lead to the desired performance. In practice, an exact output predictor would require knowledge of future disturbances and an exact model of the plant dynamics. As future disturbances are assumed to be due to a correlated white noise source, then a prediction can be made that is optimal in the least error variance sense. This form of predictor requires a model of the plant and noise dynamics so that the effects of past controls and disturbances can be predicted.

The self tuning control theory investigated is an adaption of the minimum variance controller due to Astrom et al (5). These types of self tuning control are relatively simple when compared to other forms of optimal or suboptimal adaptive control, as complex mathematical manipulation is avoided by minimising special forms of cost function containing only input/output data. The convergent self tuning control law, for a linear system disturbed by a correlated white

noise source, has been interpreted as a conventional forward path control combined with a self tuning least squares output predictor. The choice of the forward path control is made by specifying the cost function. This choice is simplified as the predictor desensitizes the system response to control and parameter changes, and as it is self tuning then the response will not be degraded in the long term by slow variation of the plant parameters. Besides the parameters that would be specified for a linear discrete control law, the self tuning control requires an estimate of time delay.

Early developments of the self tuning control (15), sacrificed considerable flexibility in choice of cost function for the ability to estimate the predictor parameters without bias when using recursive least squares techniques. If the disturbances on the plant can be modelled by a correlated noise sequence and a wider choice of cost function is required, then extended least squares estimation can be used to obtain unbiased estimates. However, in practice the significant disturbances on a HVAC plant are not noise-like, thus extended least squares is unlikely to lead to unbiased parameters. Furthermore, despite using recursive least squares estimation and the resultant biased parameters, the results of simulations and practical experiments showed that the self tuning control performnce was not noticeably degraded by bias, and thus a more complex estimation procedure was unnecessary.

The self tuning controllers investigated in this report are based on the k-incremental predictor due to Clarke et al (21), in which the data used by the least squares estimator is the k difference of the measured input/output data. As the k difference of a constant is zero then the prediction offset for this prediction is zero despite biased parameters. To reduce the steady state error (from setvalue) to zero, the cost function is chosen to give a forward path controller in the form of a discrete PI control. Besides the specification of the cost function there are several parameters that must be set before the self tuner can operate. The parameters that require specification at time of commissioning are the sample time and time delay, the other parameters are sufficiently noncritical to be given a fixed value despite the application.

The sample time is chosen in relation to the plant break frequency from an estimate of the plant storage time. This can be obtained from the process reaction curve or, as the system response will be unaffected by wide range of sample times, then an estimate of the fill time of the plant can be used. This parameter must also be specified for a conventional discrete controller.

The selection of the correct time delay is critical, because if it is sufficiently underestimated then the system response will become unstable. As this parameter is not easily estimated analytically, it must be measured from experimental data. This can be

carried out during the initial transient at plant start up, and as the delay is determined largely by the air supply rate, which is constant, then it only needs to be measured infrequently. The time delay is estimated as the time taken for the first control signal to affect the system output, which is a simple measurement to implement within the controller itself, and thus would not require to be input during commissioning.

The results of temperature responses to step changes in setvalue, showed that the initial self tuning control response is degraded due to the poor initial parameter estimates. Further setvalue change responses were not degraded as the parameter estimates had improved considerably during the first 50 samples. As the parameter values are difficult to estimate manually and vary from one plant to another, then a "CAUTIOUS" form of control used during the initial transient is likely to be the best way of avoiding such effects. Alternatively, as this degradation of transient response occurs only once, during commissioning, then it may be ignored, and no cautious control implemented. The initial parameter estimates are all set to zero except '$\hat{g}_0$', thus initially the predictor merely feeds back the measured value of the system output. It is convenient to set '$\hat{g}_0$' as the threshold value on variation in system output used when the time delay is calculated.

The number of estimated parameters can be treated liberally, as shown in chapter 5.2, because reducing the

parameter number from 10 to 4 does not result in a significant degradation of the performance. Whereas reducing the estimated time delay from 7 to 4 sample periods resulted in an increased sensitivity of the response to the variation of plant gain. However, despite this underestimation the self tuning control was shown still to be less sensitive to gain changes than a conventional PI control under the same experimental conditions.

Therefore in estimating time delay a realistic error is allowable, thus a simple estimation algorithm is likely to be adequate. Furthermore, as the number of parameters estimated can be greatly reduced without performance degradation, then it is possible to ease the storage requirements and reduce the computation time of the algorithm.

The recursive least squares algorithm is able to track slowly time varying parameters by using a forgetting factor less than unity. If the data supplied to the parameter estimator is insufficiently active, then the action of the forgetting factor may cause the magnitude of the covariance matrix elements to increase to very large values. When the data becomes active due to disturbances or a setvalue change, then the large covariance matrix will cause the parameter estimates to vary wildly, leading to a degradation of the response. To avoid such occurrences, a variable forgetting factor has been used based on an algorithm by Fortescue et al

(35). The forgetting factor is varied in relation to a simple measure of the information content of the measured system output, such that when the information content is low the forgetting factor is unity and hence the covariance matrix is not increased.

The speed of convergence of the parameter estimates can be increased if the covariance is increased, thus this algorithm can be used to obtain estimates more quickly. Also, resetting the covariance matrix to a large value identity matrix will have a similar effect. However, although the convergence may be faster, the estimates vary wildly due to increased covariance, and as demonstrated in chapter 4.5 give rise to a more oscillatory transient response to a setvalue change. As such transients are undesirable the forgetting factor algorithm was implemented such that wild variation of parameter estimates does not occur. This algorithm is part of the "jacketing" software that is required to implement self tuning control in a real application. The need for this particular algorithm arises due to the nature of the disturbances on HVAC plant.

The commissioning of a controller entails the adjustment/ setting-up of the device at the site where it is to be used, such that the transient and long term performance conforms to that specified. At the most fundamental level, this would mean for a discrete PI controller, that the sample time, the proportional gain, and the integral time are input via the operator console/keypad. The

self tuning control investigated also requires these inputs, other parameters such as the number of predictor parameters are specified at the factory, or in the case of the time delay are measured by the controller itself.

An analytical commissioning procedure was described in chapter 5.1 to allow adjustment of the PI terms for the conventional and self tuning control. Such a process would not be performed in practice, the setting up would be based on the commissioning engineer's experience, and in certain cases where the control performance is unacceptable, then the commissioning engineer may revisit the site to make fine adjustment to the controller. Various ways of commissioning the self tuning controller have been described in chapter 5.1.3 all of which require a different degree of human intervention. The exact level of desired intervention is a subject of current research. However, based on experience of the introduction of new developments in control design into the HVAC industry it is evident that some degree of customer intevention is desirable as a means of reducing the cost of warranty claims, and improving the performance of the controls. A suitable parameter to have as adjustable by commissioning engineer or customer is the proportional gain of a PI control. This is because:

(i) Adjustment of two parameters (P,I) is likely to be too difficult and the performance may be degraded if inappropriate values are used.

(ii) For the control design procedure described in chapter 5.1.2 the adjustment of control gain is directly linked with speed of response and variance of control signal. Thus it is relatively simple to relate controller gain variation to system response, and thus make efficient trial and error judgements.

(iii) The self tuning predictor makes the system response less sensitive to gain changes than a conventional controller, thus the setting of the control gain is not critical for adequate performance.

Due to the nonlinear characteristics of valves and heat exchangers, the small signal gain of the heating system model is operating point dependent. The other thermal characteristics of the heating system model are relatively constant, or do not have such a marked effect on the system response as the forward gain, so this parameter is varied to illustrate the sensitivity of the commissioned response to plant parameter changes, or to inaccuracies in the commissioning of the controller.

It is not appropriate to compare the ease of commissioning for conventional and self tuning control, as no attempt was made to find the "best" way to select

the controller parameters. Instead, the ease of control gain setting was inferred by observing the sensitivity of the commissioned response to forward gain changes. The controls were commissioned for a particular setvalue change, and thus a particular plant gain change, as the gain is operating point dependent. The response for a different setvalue change showed that the conventional control response was significantly degraded, whereas the self tuning control response was similar to that obtained during commissioning, and thus perfectly acceptable.

Thus at the most fundamental level, the self tuning control can be implemented as a conventional control except that, for plant containing significant time delay, the commissioned response is less sensitive than for conventional control.

The role of the temperature controller is mainly as a regulator, the acquisition of comfort conditions occurs most often at plant start up when the building is unoccupied. Therefore, the control response of most interest is that due to the likely disturbances to the plant. The disturbance most difficult to regulate is the virtually instantaneous change in heat stored in the space due to solar radiation or a large change in the level of occupancy. This type of disturbance was applied by step changes in measured temperature for simulated results, and by switching on a 60 Watt

incandescent lamp within the air space for experimental results. So that the sensitivity of the disturbance response of the conventional and self tuning control could be compared, the forward plant gain is varied slowly over a large time period while the disturbances are applied. For a nonlinear valve and heat exchanger characteristic this type of gain variation is due to a slowly time varying operating point as a result of the time varying mean value of the disturbance. This type of operating point variation occurs particularly in storage heating systems, where the temperature of the heat store reduces continuously over the entire occupancy period, and the operating point subsequently increases.

The time varying mean disturbance value is easily simulated, but for experimental results would require control of the ambient temperature which is not possible with this test facility. So to simulate the slowly time varying operating point dependent gain, a constant operating point and a time varying control signal gain was used.

The results of simulation and experiment showed in chapter 5.2 that the initially acceptable conventional control performance became more oscillatory as the forward gain increases and eventually became unstable in one instance, whereas the self tuning control performance was not noticeably affected. This illustrates the ability

of the self tuning control to preserve the initially commissioned performance despite slow variation in the plant parameters.

Besides the operating point dependent gain the HVAC plant exhibit nonlinearities due to valve saturation and actuator hysterisis that, in principle, can present major problems in applying self tuning control (94). The effect of valve saturation can be considered to appear to the parameter estimator as a zero gain plant, because the control signals have no effect on the system output. However due to the slow variation of the estimated parameters and the variation in the system output due to disturbances, the parameter estimates were not biased sufficiently to cause an unacceptable response once saturation ceased. Hence for the experimental conditions applied, the self tuning control performance was largely unaffected by valve saturation.

The effect of actuator hysterisis was shown in chapter 6.2 to be potentially very detrimental to the performance of self tuning or conventional control. If the hysterisis is known then compensation can be devised to reduce its effects, however as in practice it is unknown, then the most common means of reducing the oscillation it causes is to reduce the control gain. For a typical worst case value of hysterisis and a linear valve and heat exchanger characteristic, the response of the self tuning controller was largely unaffected.

However, if in the event of a particularly nonlinear valve causing a high value of gain coupled with a significant value of hysterisis, then the response is likely to be degraded. In such an extreme instance the most convenient course of action would be to reduce the control gain, whether it be a conventional or self tuning controller. This is then an example of where a control gain that is user adjustable would be an advantage, as it could be used to quickly remedy the hysterisis effects.

No attempt has been made to simplify or compress the algorithms to save on data and program storage. However, several aspects of the basic self tuner are suitable for simplification, and the algorithm has been shown to be amenable to certain simplifications, such as reduced estimator parameter number.

The self tuning control lends itself to two main implementations:

(i) As a stand-alone controller

(ii) As one of several control algorithms in the same package, sharing computational facilities

The first implementation would require a code efficient algorithm, thus the full recursive least squares algorithm would have to be simplified, a fixed point number representation further reducing the computational load. Research (15,28) has shown that these simplifications

are feasible, allowing implementation of the self tuning controller on a single chip microcontroller device. However, due to the simplification of the parameter estimator the speed of convergence of the estimates is likely to be reduced, thus some degradation of performance is inevitable.

The second implementation may be a more realistic proposition for this particular application. Due to the large sample periods common for air temperature control, a single microprocessor could be used to carry out the necessary computation to implement several self tuning control loops. As all the self tuning controls would require an algorithm to update parameter estimates, then this could be shared by all the control algorithms. Similarly, the scaling problems and loss of resolution due to fixed point number representation can be avoided by using a floating point representation. The increased cost of developing the software, data/program storage, and data input/output facilties being offset by the savings in implementing several controllers without the need to reproduce the hardware several times. In an effort to reduce the cost of developing software , it is becoming more popular to write control programs in medium level languages such as PLM 80,51. However as the least squares algorithm is likely to be used unaltered for a considerable time, then it may be cost effective to write this algorithm in assembler code.

The self tuning temperature controller, if it is to be competitive, must be code efficient so that the controller can offer the same range of facilities as the conventional controller, at a similar cost. It is not sufficient that a cost saving will be made in commissioning and warranty claims, the cost of producing the controller must also be similar before manufacturers will accept self tuning control. No attempt has been made to optimise the storage requirements or computation time for the algorithm described, however the key work by Clarke et al (15) which implements a self tuning controller using an 8080 microprocessor, does give some indication of the requirements for a well structured algorithm. It is obvious that the requirements for the PI control cannot be compared to that for the self tuner, because besides the PI-type cost function, the predictor and parameter estimator calculations must be carried out. However, comparison can be made with more complex conventional controls, such as the 3 stage PID control algorithm. This controller has provision for the proper sequencing of a 3 stages of heating, cooling, or dampers, or a combination of these. The complexity of this algorithm is increased due to the necessary filtering of the derivative function, wind-up compensation, and the measures taken to ensure a smooth transition between stages. The storage requirements for the self tuning controller as given by Clarke are as follows,

program code approx. 2½k bytes
data approx. ½k bytes

A similar implementation of the PID algorithm has storage requirements as follows,

program code approx. 2k bytes

data approx. ¼k bytes

Thus it is possible to implent self tuning control laws using similar data and program storage capacity as needed for certain conventional control algorithms. The original self tuning algorithm took approximately 800mS to update 10 parameters, when operating with a 2MHz clock frequency. It is now common for industry standard microprocessors to operate on a 4MHz clock frequency, so 10 separate control signals can be calculated in approximately 4 seconds.

As part of the author's work in industry an Energy Management Controller was designed that implements 3,3 stage PID controllers, 3 self commissiong controllers, 3 plant optimisers, and extensive temperature, humidity and alarm monitoring. It is envisaged that as more data is obtained of self-tuning control of real plant, and confidence of its long term performance is had, then the self commissioning control algorithms will be replaced by self tuning algorithms.

8. Further Work

During the course of this research there have been several aspects of self tuning control and the control of heating plant in particular that invite further study.

8.1 Heating plant control

The level of manual commissioning performed by the commissioning engineer should ideally be kept to a minimum, so that the total cost of the control equipment and its installation can be minimised. Furthermore, if the long term performance of the control is unsatisfactory then the commissioning engineer may have to pay a visit to the site to readjust the control parameters, and the cost of the visit may be claimed against the manufacturers warranty agreement. This would then result in an overall increase in the cost of the controller, thus site visits must be minimised. To reduce the likely cost of the controller it may be wise to allow the customer some influence over the control performance. Thus in the event of short or long term unsatisfactory performance, the customer may make what adjustments are available, and if these are insufficient then a site visit will be necessary. The self tuning control previously described allowed a user definable control signal cost function. Increasing the gain of the cost function increases the control weighting and subsequently reduces the control signal variance. Thus the customer can redefine the control signal variance,

which may be too great due to hysterisis or plant nonlinearity.

This discussion raises two major topics that invite further study:

(i) What level of customer intervention is desired, or is possible?

(ii) What advantages are there in a user definable control variance setting compared to a conventional proportional band setting?

At the start and end of occupancy the boiler will be switched off/on by a timeswitch or energy optimiser. The savings accrued by using an energy optimiser is so significant that most large buildings use some form of plant operating period optimisation. A microprocessor based optimiser will have an algorithm that uses a model of the thermal characteristics of the building inputs such as inside and outside air temperature, to predict the time it will take the room temperature to change by a given amount. For simplicity it is commonly assumed that the valve position is unaltered during the preheat time, which by necessity precludes closed loop control of air temperature. Hence, due to the interaction between control and optimisation, errors in preheat and precool calculations may arise, leading to energy wastage, and hence this subject invites further research.

8.2 Self tuning control

The discussion of the implementation of the self tuning controller has primarily considered an integrated controller, implementing several control loops using one main microprocessor. The advantages of this approach being that due to the reduced cost of the hardware for each loop, and as the control algorithms can share the common routines, then it is feasible to use the full algorithms without simplification. Furthermore, the programs can be written in a medium or high level language, thus simplifying the software development. However, in several situations a single loop controller is all that is required. This is especially relevant where part of an existing air conditioning system requires to be upgraded or modified, for instance the addition of a computer suite, conference room, or operating theatre to an existing building. For such controllers, the program storage capacity and computational power of the microprocessor is limited by cost. This is so that the controller can remain competitive with conventional products such as PI or PID controllers. The major computational and storage requirements of the self tuning controller is that due to the least squares parameter estimator, and the arithmetic functions. Various simplifications of the least squares algorithms are possible, however it is most common to attempt modifications of the covariance matrix calculation (15,28). As floating point arithmetic functions are very code consuming then fixed point routines are

preferable. However, the necessary scaling to ensure the most significant digits are retained, whilst the desired resolution is achieved is a particularly difficult problem. Investigation of the simplifications to the self tuning control algorithm, in an effort to optimise the computation and storage load, is a complex subject and deserves further study.

The major commissioning responsibility is the definition of the cost function, all other parameters are preset, or can easily be set by the controller based on data obtained from the plant. Once it has been established what level of user intervention is desirable, then the commissioning process can be designed. Previous study of this subject (29) has been based on a deterministic search algorithm to obtain the cost function. Although this algorithm estimates all the cost function parameters, it could be modified so that only those parameters that are difficult to set, are set automatically, and other parameters such as control signal weighting are set by the user. Another approach to this problem is to identify characteristics of the heating plant from input/output data and use a procedure such as that described in chapter 5.2 to select appropriate values for the cost function. This scheme can be termed a "hybrid self tuner" as it combines the implicit approach such as that described by Clarke (15), and the explicit pole-placement approach as described by Wellstead et al (89). This subject requires further research combined

with a subjective investigation of the necessary and desirable level of user intervention.

REFERENCES

1. Allidina, A. Y., Hughes, F. M.,
"Self tuning controller with integral action",
UMIST Control Systems Center Report No. 526,
Sept. (1981).

2. Åström, K. J.,
"Introduction to stochastic control theory",
Academic Press, New York and London, (1971)

3. Åström, K. J., Eykhoff, P.,
"System identification - A survey"
Automatica, Vol. 7, pp 123-162, (1971)

4. Åström, K. J., Wittenmark, B.,
"On self tuning regulators",
Automatica, Vol. 9, pp 185-189, (1973)

5. Åström, K. J., Wittenmark, B., Borrison, U.,
Ljung, L.,
"Theory and applications of self tuning regulators",
Automatica, Vol. 13, pp 457-476, (1977)

6. Balasubramarian, S., Rajagopalan, T.,
"Compensation of dead time effect by anticipatory
controller in thermo-electric control system"
Proc. of the 2nd Int. Conf. on Thermoelectric
Energy Conservation, Arlington USA, pp 48-53,
(1978)

7. Bergman, M. S.,
"Application of incremental control to HVAC
processes",
Proc. of the 1982 American Control Conf.,
Arlington USA, pp 1133-1139, (1982).

8. Billington, N. S.,
"Building physics: Heat"
The Commonwealth and International Library of
Science Technology and Liberal Studies, (1967)

9. Building Research Establishment,
"Energy conservation: a study of energy
consumption in buildings and possible means of
saving energy in houses",
"Energy, Heating and Thermal Comfort"
The Construction Press, Lancaster, England (1978)

10. Changse, K., Friedly, J. C.,
"Approximate dynamic modelling of large staged
systems",
Industrial Engineering Chemistry, Process Design
and Development, Vol. 13, No. 2, pp 177-181, (1974)

11. Chiang, H. S., Durbin, L. D.,
"Gain-adaptive control applied to a heat exchange process using a first order plus dead time compensator.

12. Clarke, D. W.,
"Generalised-least-squares estimation of the parameters of a dynamic model",
IFAC Symposium of Idnetification, Prague, (1967)

13. Clarke, D. W., Briggs, P. A. N.,
"Errors in weighting sequence estimation. II The effects of autocorrelated noise",
Int. J. Cont., Vol. 11, No. 1, pp 57-65, (1970)

14. Clarke, D. W., Gawthrop, P. J.,
"Self-tuning controller",
Proc: IEE, Vol. 22, No. 9, pp 929-934, (1975).

15. Clarke, D. W., Cope, S. N., Gawthrop, P. J.,
"Feasibility study of the application of microprocessors to self-tuning controllers"
OUEL Report No. 1137/75 (1975)

16. Clarke, D. W., Frost, P. J.,
"A Control Basic for microcomputers",
IEE Conf. on Trends in Online Computer Control Systems.

17. Clarke, D. W., Gawthrop, P. J.,
"Self tuning control"
Proc. IEE, Vol. 126, No. 6, pp 633-640, (1979)

18. Clarke, D. W.,
"Introduction to self tuning controllers",
Self Tuning and Adaptive Control: Theory and Applications, Peter Peregrinus, (1981)

19. Clarke, D. W.,
"Implementation of self tuning controllers"
ibid

20. Clarke, D. W., Gawthrop, P. J.,
"Implementation and application of microprocessor based self-tuners",
Automatica, Vol. 17, No. 1, pp 233-244, (1981)

21. Clarke, D. W., Hodgson, A. J. F., Tuffs, P. S.,
"Offset problem and k-incremental predictors in self-tuning control",
Proc. IEE, Vol. 130, Pt.D, No. 5, pp 217-225, (1983)

22. Cope, S. N.,
"Floating point arithmetic routines and macros for an Intel 8080 microprocessor"
OUEL Report No. 1123/75, (1975)

23. Croome, D. J.,
"Man, environment and buildings",
Building and Environment, Vol. 15, pp 235-238,
(1980)

24. Dahlin, E. B., Zeimer, R. L., Wickstrom, W. A.,
Horner, M. G.,
"Designing and tuning digital controllers"
Instruments and Control Systems USA, July,
pp 87-91, (1968)

25. Dexter, A. L.,
"Self tuning optimum start control of heating
plant",
Automatica, Vol. 17, No. 3, pp 483-492, (1981)

26. Dexter, A. L.,
"Self tuning charge control scheme for domestic
stored-energy heating systems", Proc.IEE,
Vol. 128, Pt.D, No. 6, pp 292-300, (1981)

27. Dexter, A. L.,
"Self adaptive control of hot water space heating
systems",
Proc. of IASTED Symposium EES 83 on Energy and
Environmental Systems, Athens, (1983)

28. Dexter, A. L.,
"Self tuning control algorithm for single chip
microcomputer implementation",
Proc IEE, Vol. 130, Pt.D, No. 5, (1983)

29. Dexter, A. L., Danninger, W., Graham, W. J.,
"Design of self tuning controllers for wet
heating systems",
OUEL Report No. 1531/84, (1984)

30. Fan, L. T., Murang, Y. S., Murang, C. L.,
"Applications of modern optimal control theory
to environmental control of confined spaces and
life support systems",
Build. Sci., Vol. 5, pp 57-71, (1970)

31. Farris, D. R., McDonald, T. E.,
"Adaptive optimal control algorithm for DDC",
Trans. ASHRAE, Pt 1, pp 880-893, (1980).

32. Fertik, H. A., Ross, C. W.,
"Direct digital control algorithm with anti
windup feature",
Trans. ISA, Vol. 6, No. 4, pp 317-328, (1967)

33. Fertik, H. A.,
"Tuning controllers for noisey processes",
Trans. ISA, Vol. 14, pp 292-304, (1975)

34. Fisk, D. J.,
"Mean square error as a criterion for temperature control",
Building Services Engineering Research and Technology, Vol. 2, No. 3, pp 127-132, (1981).

35. Fortescue, T. R., Kershenbaum, L. S., Ydstie, B. E.,
"Implementation of self tuning regulators with variable forgetting factors",
Automatica, Vol. 17, No. 6, pp 831-835, (1981)

36. Franklin, G. F., Powell, J. D.,
"Digital control of dynamic systems",
Addison-Wesley, Phillipines, (1980)

37. Gallier, P. W., Otto, R. E.,
"Self tuning computer adapts DDC algorithms",
Instrumentation Technnol., Vol. 15, No. 2, pp 65-70, (1968)

38. Gawthrop, P. J.,
"Some interpretations of the self tuning controller",
Proc. IEE, Vol.124, No. 10, pp 889-892, (1977)

39. Giblin, J.,
"Robust self tuning control applied to heating systems",
PhD Thesis, University of Dublin, (1981)

40. Gilbaro, L. G., Lees, F. P.,
"The reduction of complex tranfer function models to simple models using the method of moments",
Chem. Eng. Sc., Vol. 24, pp 85-93, (1969)

41. Gilles, G.,
"New results in modelling heat exchanger dynamics",
Trans. ASME, Journal of Dynamic Systems, Measurement and Control, Sept., pp 277-282, (1974).

42. Godfrey, K. R.,
"The theory of the correlation method of dynamic analysis and its application to industrial processes and nuclear power plant",
Measurement and Control, Vol. 2, May, pp 765-772, (1969)

43. Godfrey, K. R., Moore, D. J.,
"Identification of processes having direction dependent responses, with gas turbine applications"
Automatica, Vol. 10, pp 469-481, (1974)

44. Goodwin, G. C., Elliott, A. B., Teoh, E. K.,
"Deterministic convergence of a self tuning regulator with covariance resetting",
Proc. IEE, Vol. 130, Pt.D, No. 1, pp 6-8, (1983)

45. Green, M. D., Ulge, A.,
"Frequency and time domain thermal response of dwellings",
Building and Environment, Vol. 14, pp 107-108, (1979)

46. Haalman, A.,
"Adjusting controllers for a deadtime process",
Control Engineering, No. 7, July, pp 87-91, (1968)

47. Hanna, G. B.,
"Computer analysis to estimate the temperature response of an enclosure by the finite difference method",
Building Services Engineer, Vol. 45, Aug., pp 59-68, (1977)

48. Harrison, H. L., Hansen, W. S., Zelenski, R. E.,
"Development of a room transfer function model for use in the study of short term transient response",
Trans. ASHRAE, Vol. 74, pp 198-210, (1968)

49. Hastings, J. R., Sage, M. W.,
"Recursive generalised least squares procedures for online identification of process parameters",
Proc. IEE, Vol. 116, No. 12, pp 2057-2062, (1969)

50. Holmes, M. J., Adams, S.,
"Saving energy by a better understanding of systems and controls",
BSRIA Report, Nov., (1976)

51. Hsia, T. C.,
"On least squares algorithms for system parameter estimation",
Trans. IEEE, Automatic Control, Feb., pp 104-108, (1976)

52. Hsia, T. C.,
"System Identification: Least Squares Methods",
Lexington Books, (1977)

53. Isermann, R.,
"Practical aspects of process identification",
Automatica, Vol. 16, pp 575-587, (1980)

54. Isermann, R.,
"Parameter adaptive control algorithms - a tutorial",
Automatica, Vol. 18, No. 5, pp 513-528, (1982)

55. Jacobs, D., Donaghey, L. F.,
"Microcomputer implementation of direct digital control algorithms for thermal process control",
Trans. ASME, Journal of Dynamic Systems Measurement and Control, Dec., pp 233-240, (1977)

56. James, R. W.,
"Modelling and control of refrigeration and air conditioning systems for energy conservation",
PhD Thesis, University of Sheffield, (1976)

57. Kalman, R. E.,
"Design of a self optimising control system",
Trans. ASME, pp 468-478, Feb., (1958)

58. Krikelis, N. J.,
"State feeback integral control with intelligent integrators",
Int. J. Cont., Vol. 32, No. 3, pp 465-473, (1980)

59. Leiper, D., Powers, I.,
"A self optimising central heating control system",
Microelectronics Reliability, Vol. 18, pp 101-109, (1978)

60. Leth erman, K. M., Palin, C. J., Park, P. M.,
"The measurement of dynamic thermal response in rooms using pseudo-random binary sequences",
Building and Environment, Vol. 17, No. 1, pp 11-16, (1982)

61. Ljung, L., Wittenmark, B.,
"asymptotic properties of self tuning regulators",
Lund Institute of Technology, Report 7404, (1974)

62. Lush, D. M.,
"Microprocessors for building services and energy conservation",
Electronics and Power, Jan., pp 75-79, (1981)

63. MacGregor, J. F., Wright, J. D., Huynh Man Hong,
"Optimal tuning of digital PID controllers using dynamic-stochastic models",
Ind. Eng. Chem., Proc. Des. and Dev., Vol. 14, No. 4, pp 398-402, (1975)

64. Marshall, J. E.,
"Control of Time Delay Systems",
Peter Peregrinus, (1971)

65. Murtagh, K. J.,
"An investigation into the use of control systems of variable structure for environmental control",
Final year report, Bath University, (1981)

66. Nakamishi, E., Pereira, N. C., Fan, L. T.,Husang, C. L., "Simultaneous control of temperature and humidity in a confined space", Building Science, Vol. 8, pp 39-49, (1973)

67. Orbach, A., Rorres, C., Fischl, R., "Optimal control of a solar collector loop using a distributed-lumped model", Automatica, Vol. 17, No. 3, pp 535-539, (1981)

68. Osorio Cordero, A., Mayne, D. Q., "Deterministic convergence of a self tuning regulator with variable forgetting factor", Proc. IEE, Vol, 128, Pt.D, No. 1, pp 19-23, (1981).

69. Palmor, Z. J., Shinnar, R., "Design and tuning of dead time compensators", Proc. Joint Auto. Cont. Conf. Philadelphia USA, pp 59-70, Oct (1978)

70. Paynter, H. M., Takahashi, Y., "A new method of evaluating dynamic response of counterflow and parallel flow heat exchangers", Trans. ASME, Vol. 78, pp 749 - , (1956)

71. Peterka, V., "A square root filter for real time multivariate regression", Kybernetika, Vol. 11, No. 1, pp , (1975)

72. Pimbert, S. L., "Warm air heating - thermostat differential and the setting of thermostats", Building Services Engineering Research and Technology, Vol. 4, No. 1, pp 1-5, (1983).

73. Radke, F., "Application of a microprocessor based controller for parameter adaptive control of an air heater", IEEE Conf. on Applications of Adaptive and Multivariable Control, Hull, pp 134-139, (1982)

74. Rolls, T., Willis, A., Gwinn, R., "Design concept for a microprocessor based temperature controller", Microprocessors and Microsystems, Vol. 6, No. 5, pp 225-233, (1982)

75. Roots, W. K., "Fundamentals of temperature control", Academic Press, (1969)

76. Ross, C. W., Green, T. A., Schwartzenberg, J. W., "Experimentally determined properties of exponentially correlated gaussian noise", Trans. ISA, Vol. 13, No. 2, pp 172-181, (1974)

77. Ross, C. W.,
"Evaluation of controllers for deadtime processes",
Trans. ISA, Vol. 16, No. 3, pp 25-34, (1977)

78. Private communication,
Technical Dept., Satchwell Control Systems Ltd,
Slough, Berks.

79. Satchwell Guide to Environmental Control,
ibid.

80. Satchwell Product Specification,
ibid.

81. Schumann, R.,
"Digital parameter adaptive control of air-conditioning plant",
Automatica, Vol. 18, No. 5, pp 569-575, (1982)

82. Simpson, H. R.,
"Statistical properties of a class of psuedorandom sequences",
Proc. IEE, Vol. 113, No. 12, pp 2075-2080, (1966)

83. Soderstrom, T.,
"On the generalised least squares method. Counter examples to general convergence",
Automatica, Vol. 10, pp 681-683, (1974)

84. Talmon, J. L., Van Den Boom, A. J. W.,
"On the estimation of the transfer function parameters of process and noise dynamics using a singlestage estimator",
3rd IFAC Sym. on Identification and System Parameter Estimation, pp 711-720, (1973)

85. Tobias, J. R.,
"Design considerations for adaptive control of buildings",
Proc. IEEE Conf. on Decision and Control, San Diego USA, PP 214-216, Jan, (1979)

86. Udink ten Cate,
"Digital adaptive control of a glasshouse heating system",
IFAC Symposium on Digital Computer Applications to Process Control, pp 505-512, (1977)

87. Udink ten Cate,
"A least-squares-like gradient method for discrete process identification",
Int. J. Control, Vol. 28, No. 6, pp 933-952, (1978)

88. Uronen, P., Yliniemi, L.,
"Experimental comparison and application of different DDC - algorithms",
IFAC Sym. on Digial Computer Applications to Process Control, pp 457-464, (1977)

89. Wellstead, P. E., Prager, D., Zanker, P.,
"Pole assignment self tuning regulator",
Proc. IEE, Vol. 126, No. 8, pp 781-787, (1979)

90. Welwyn thermistor data sheet UUA, 41J54

91. Williams, B. J., Clarke, D. W.,
"Plant modelling from PRBS experiments",
Control, October, pp 856-860, (1968) and November, pp 947-951, (1968)

92. Wolsey, W. H.,
"Basic principles of automatic control",
Hutchinson Educational, (1975)

93. Young, P. C.,
"Applying parameter estimation to dynamic systems Part 1",
Control, October, pp 119-125, (1969)

94. Zanker, P. M., Wellstead, P. E.,
"Practical features of self tuning",
IEE, Trends in Online Computer Control Systems, pp 160-164, (1979)

95. Zermuehlen, R. O.,
"Room temperature response to a sudden heat disturbance input",
Trans. ASHRAE, Vol. 71, No. 1, pp 206-211, (1965)

APPENDIX A

Derivation of least squares predictor of the auxilliary system output

This derivation is a development of that originally proposed by Aström (2).

The linear system output is given by,

$$Y(t) = Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})}.U(t)+\frac{C(Z^{-1})}{A(Z^{-1})}.\xi(t)+d \qquad (A.1)$$

where the significance of the symbols is explained in chapter 2.

Define the auxilliary output '$\phi(t)$',

$$\phi(t) = P(Z^{-1}).Y(t) \qquad (A.2)$$

where

$$P(Z^{-1}) = \frac{Pn(Z^{-1})}{Pd(Z^{-1})} \qquad (A.3)$$

where '$Pn(Z^{-1})$, $Pd(Z^{-1})$' are polynomials in the backward shift operator.

From equation A.1 and equation A.2, the auxilliary output is given by,

$$\phi(t)=\frac{Z^{-k}.P(Z^{-1}).B(Z^{-1})}{A(Z^{-1})}.U(t)+\frac{P(Z^{-1}).C(Z^{-1}).\xi(t)}{A(Z^{-1})}+P(Z^{-1}).d \qquad (A.4)$$

Thus the k-step-ahead auxilliary output is given by,

$$\phi(t+k)=\frac{P(Z^{-1}).B(Z^{-1})}{A(Z^{-1})}.U(t)+Z^{k}.\frac{P(Z^{-1}).C(Z^{-1})}{A(Z^{-1})}.\xi(t)+Z^{k}.P(Z^{-1}).d \qquad (A.5)$$

The noise term can be separated into future and other terms, that is

$$Z^{k}.\frac{P(Z^{-1}).C(Z^{-1})}{A(Z^{-1})}.\xi(t)=e_0.\xi(t+k)+e_1.\xi(t+k-1)+.....$$

$$+e_{k-1}.\xi(t+1)+e_k.\xi(t)+e_{k+1}.\xi(t-1)+ ..... \qquad (A.6)$$

This can be expressed in closed form using the Diophantine as follows,

$$\frac{C(Z^{-1}).Pn(Z^{-1})}{A(Z^{-1}).Pd(Z^{-1})}=E(Z^{-1})+\frac{Z^{-k}.F(Z^{-1})}{A(Z^{-1}).Pd(Z^{-1})} \qquad (A.7)$$

where

$$E(Z^{-1})=e_0+e_1.Z^{-1}+ ..... + e_{k-1}.Z^{-(k-1)} \qquad (A.8)$$

$$F(Z^{-1})=f_0+f_1.Z^{-1}+ ..... + f_{n+m-1}.Z^{-(n+m-1)}$$

where 'm' is the order of '$Pd(Z^{-1})$'. From equation A.7 and equation A.5.

$$\phi(t+k)=\frac{P(Z^{-1}).B(Z^{-1})}{A(Z^{-1})}.U(t)+Z^{k}.E(Z^{-1}).\xi(t)+\frac{F(Z^{-1})}{A(Z^{-1}).Pd(Z^{-1})}.\xi(t)$$

$$+Z^{k}.P(Z^{-1}).d \qquad (A.9)$$

From equation A.4 the present value of the noise is given by,

$$\xi(t)=\frac{A(Z^{-1})}{P(Z^{-1}).C(Z^{-1})}.\phi(t)-Z^{-k}.\frac{B(Z^{-1})}{A(Z^{-1})}.U(t)-\frac{A(Z^{-1})}{C(Z^{-1})}.d \qquad (A.10)$$

Using equation A.10 eliminate the present noise from equation A.9, that is,

$$\phi(t+k)=\frac{P(Z^{-1}).B(Z^{-1})}{A(Z^{-1})}.U(t)+Z^{k}.E(Z^{-1}).\xi(t)+\frac{F(Z^{-1})}{Pn(Z^{-1}).C(Z^{-1})}.\phi(t)$$

$$-Z^{-k}.\frac{F(Z^{-1}).B(Z^{-1})}{A(Z^{-1}).Pd(Z^{-1}).C(Z^{-1})}.U(t)-\frac{F(Z^{-1})}{Pn(Z^{-1}).C(Z^{-1})}.d$$

$$+ Z^{k}.P(Z^{-1}).d \qquad (A.11)$$

therefore,

$$\phi(t+k)=Z^{k}.E(Z^{-1}).\xi(t)+\frac{F(Z^{-1}).\phi(t)}{Pn(Z^{-1}).C(Z^{-1})} +$$

$$\left\{\frac{P(Z^{-1}).C(Z^{-1})}{A(Z^{-1})} - \frac{Z^{-k}.F(Z^{-1})}{A(Z^{-1}).Pd(Z^{-1})}\right\}.\left\{\frac{B(Z^{-1})}{C(Z^{-1})}.U(t)+\frac{Z^{k}.A(Z^{-1})}{C(Z^{-1})}.d\right\}$$

(A.12)

Recall equation A.7, substituting in A.12 gives,

$$\phi(t+k)=Z^{k}.E(Z^{-1}).\xi(t)+\frac{F(Z^{-1}).}{Pn(Z^{-1}).C(Z^{-1})}\phi(t)+\frac{B(Z^{-1}).E(Z^{-1})}{C(Z^{-1})}.U(t)$$

$$+ \frac{Z^{k}.A(Z^{-1}).E(Z^{-1})}{C(Z^{-1})}.d \qquad (A.13)$$

The RHS of equation A.13 contains a term made up of future disturbances which are unknown. However, as the disturbances are uncorrelated and zero mean, then the best guess of their value is their mean value of zero.

The optimal predictor 'ϕ^{*}' is then given by setting the future value of the noise terms to zero (2). That is,

$$C(Z^{-1}).\phi^{*}(t+k|t)=\frac{F(Z^{-1})}{Pn(Z^{-1})}.\phi(t)+G'(Z^{-1}).U(t)+\delta \qquad (A.14)$$

where

$$G'(Z^{-1}) = E(Z^{-1}).B(Z^{-1}) \qquad (A.15)$$

$$\delta = A(1).E(1).d$$

as 'd' is a constant.

For the special case,

$$Pn(Z^{-1}) = Pd(Z^{-1}) = 1 \qquad (A.16)$$

then the optimal predicted output is given by,

$$C(Z^{-1}).Y^*(t+k|t)=F(Z^{-1}).Y(t)+G'(Z^{-1}).U(t)+\delta \qquad (A.17)$$

Note "F,G'" are not the same as in equation A.14 because from equation A.7 the value of 'F,E' are dependent on P, and 'G'' is dependent on 'E' from equation A.15.

The prediction of equation A.17 is that which gives least error variance for the system and noise described by equation A.1.

Fig. 1.1.1 Constant volume mixed air system

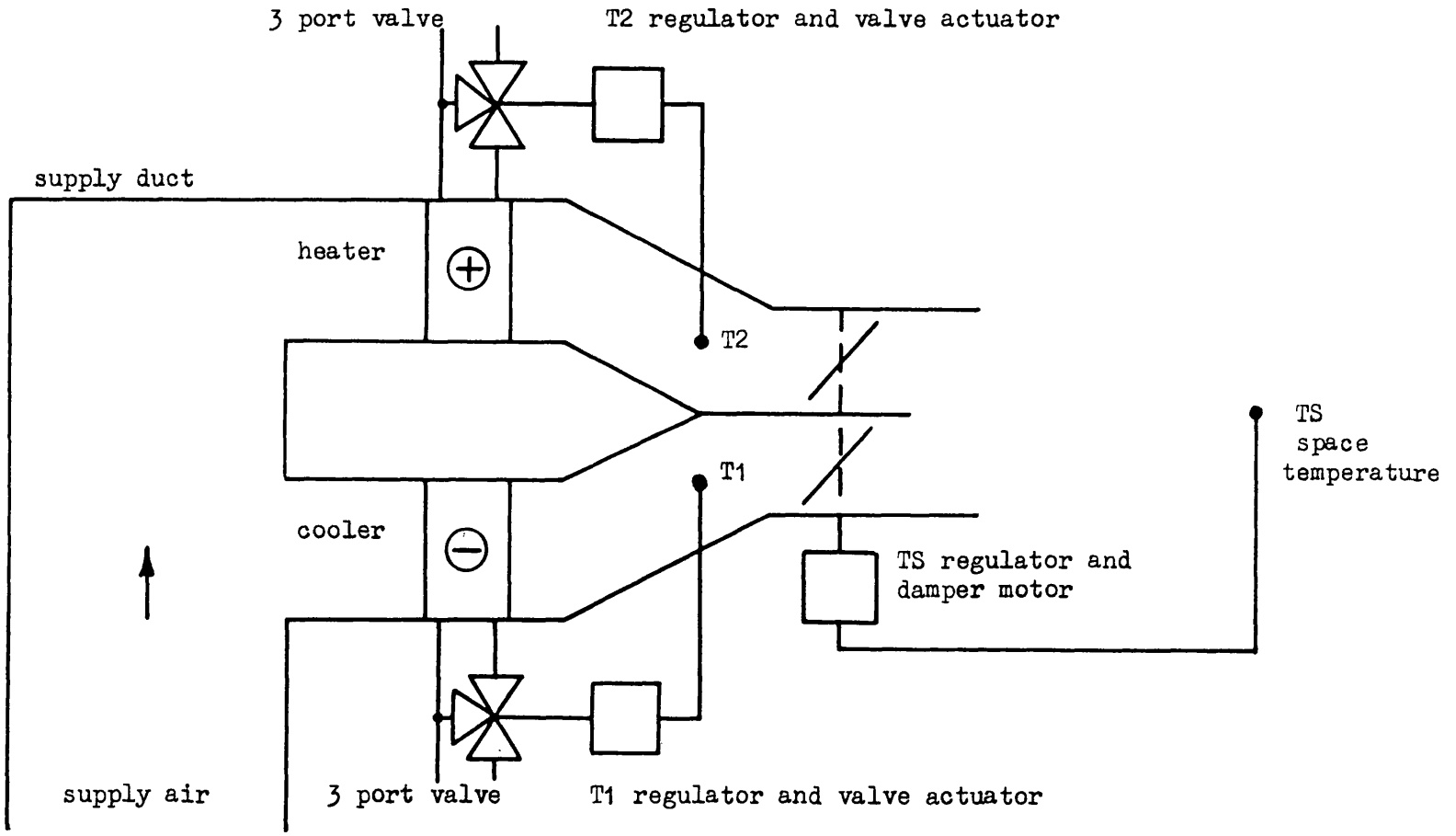


Fig. 1.1.2 Terminal reheat system

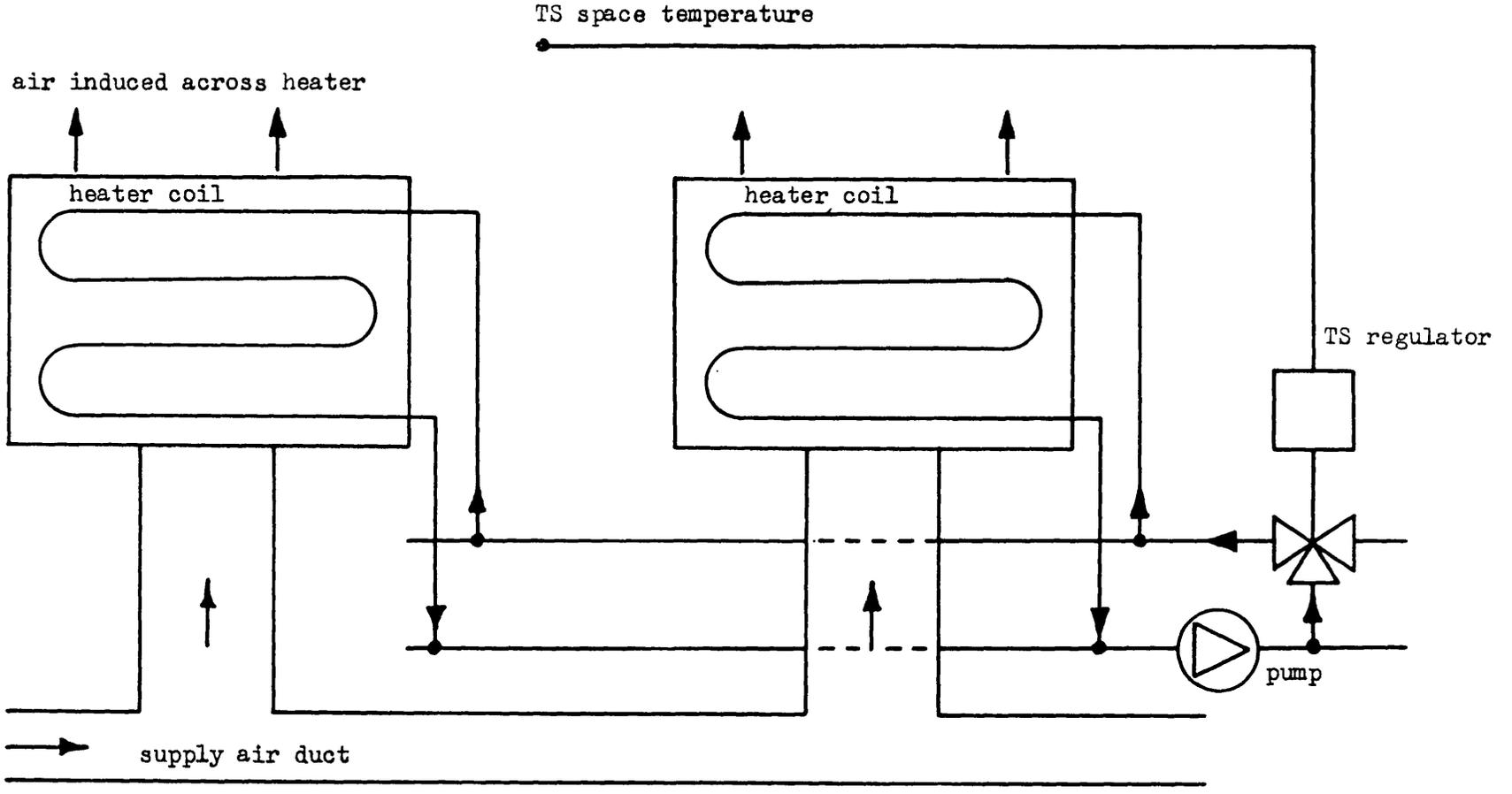


Fig. 1.1.3 3 stage , heating , cooling , and damper system

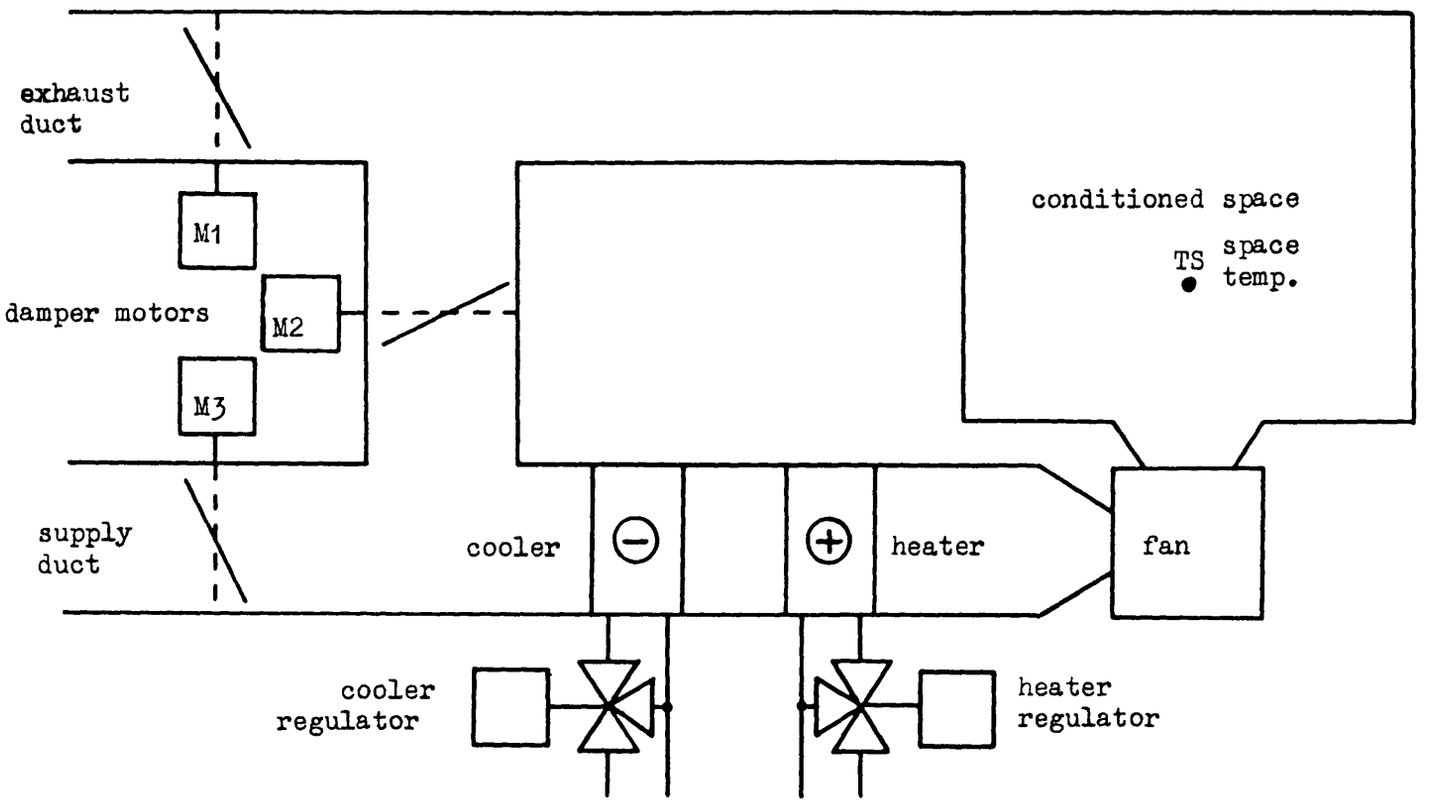
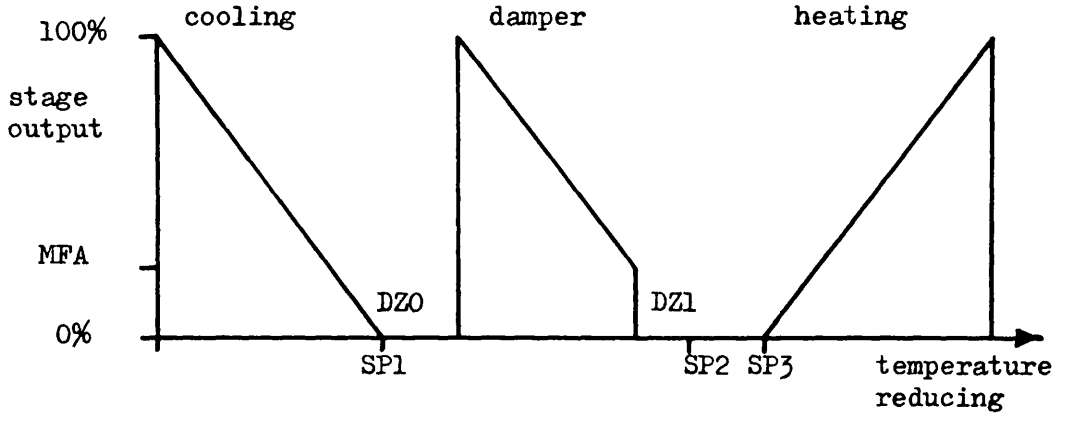


Fig. 1.1.4 Operation of stages with respect to temperature for proportional control



SP setpoint
DZ deadzone
MFA minimum fresh air setting

Fig. 1.1.5 Pressure drops in a 2 port valve water circuit

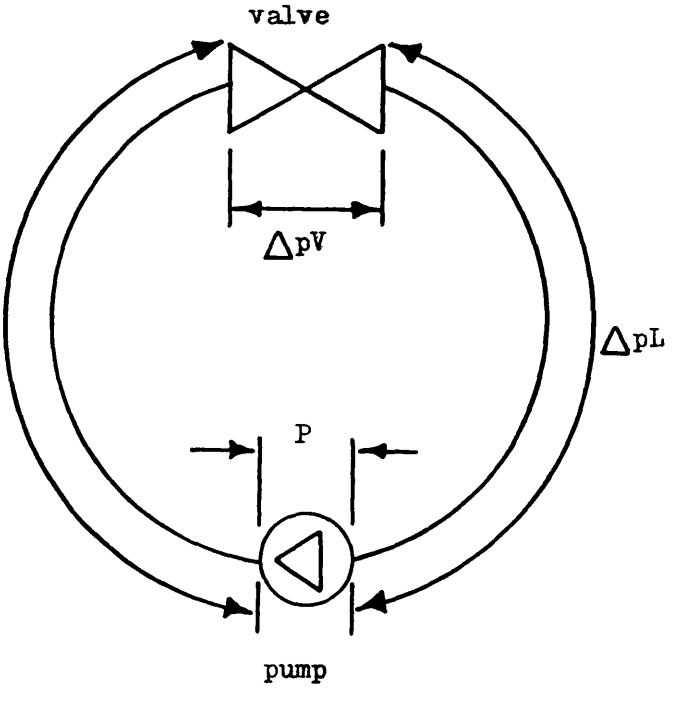


Fig. 1.1.6 Linear valve characteristic for varying authority

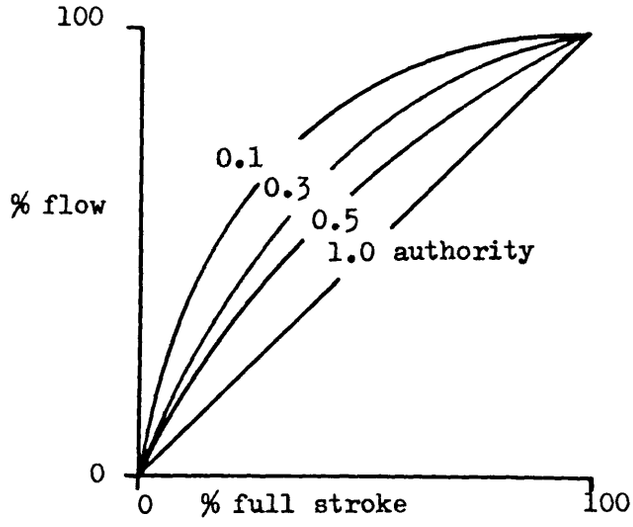


Fig. 1.1.7 Heat emission with respect to primary supply flow

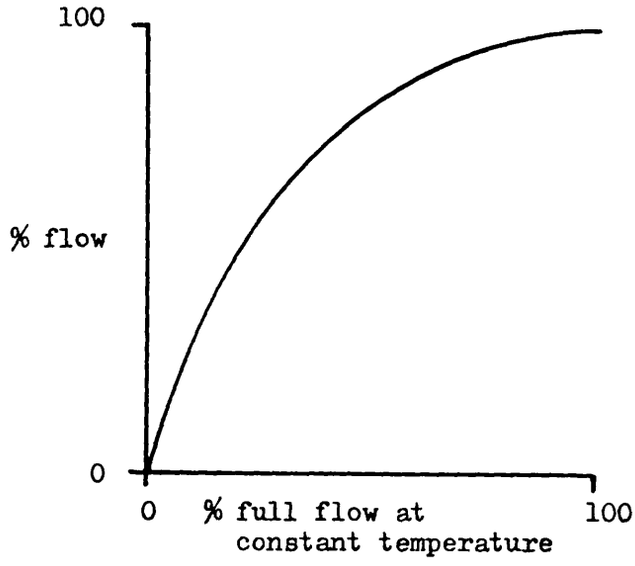


Fig. 1.1.8 Primary supply flow with respect to valve stroke for various valve geometries / authorities

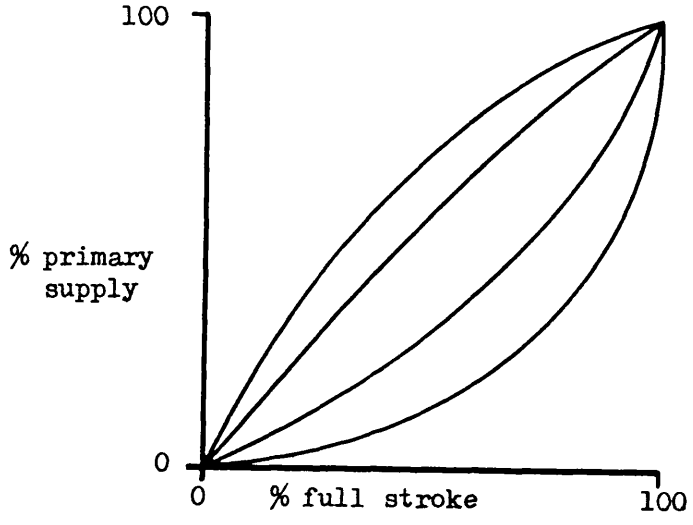


Fig. 1.1.9 Heat emission with respect to valve stroke

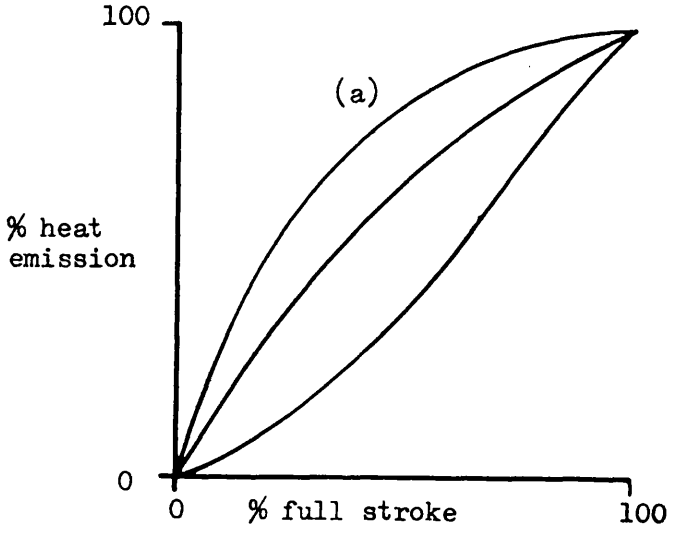


Fig. 1.2.1 Laboratory ambient air temperature for a one week period.

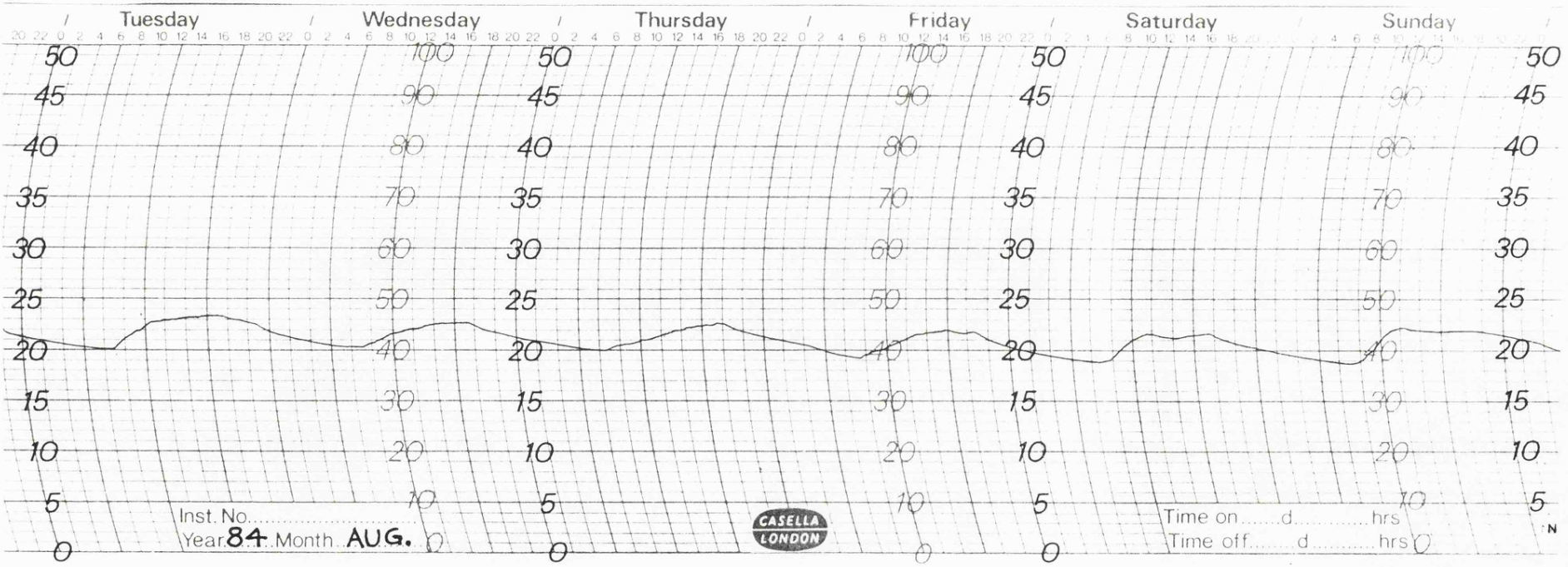


Fig. 1.2.2 Schematic diagram of section through test facility air space

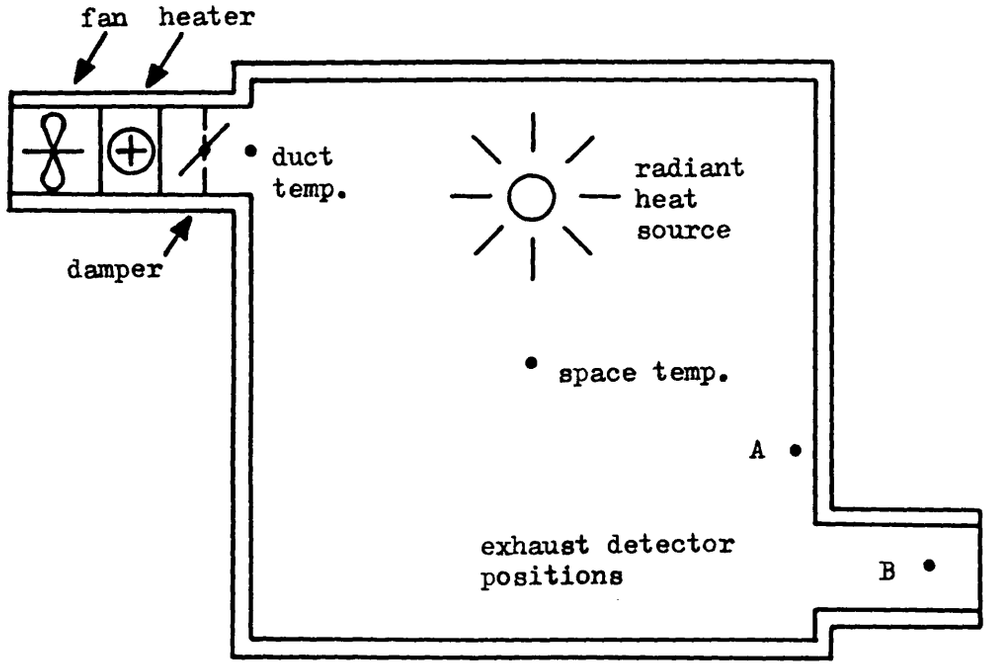


Fig. 1.2.3.a Zero crossing detector

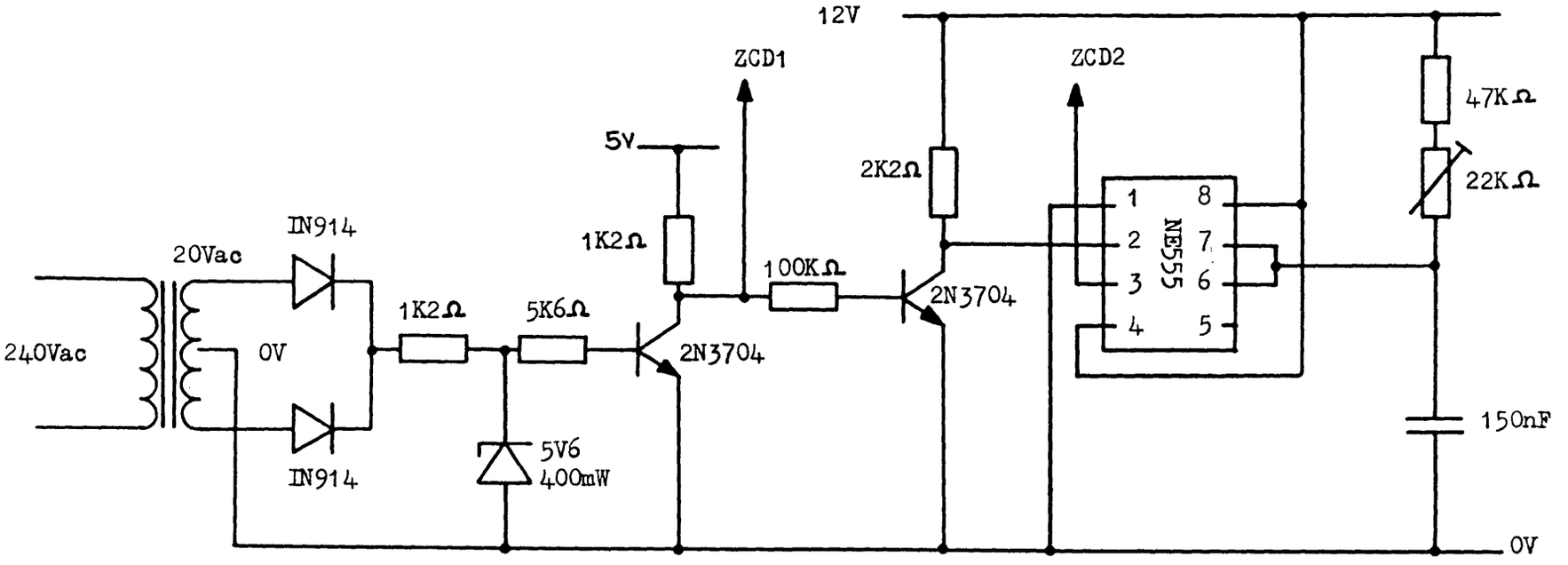


Fig. 1.2.3.b Phase locked loop

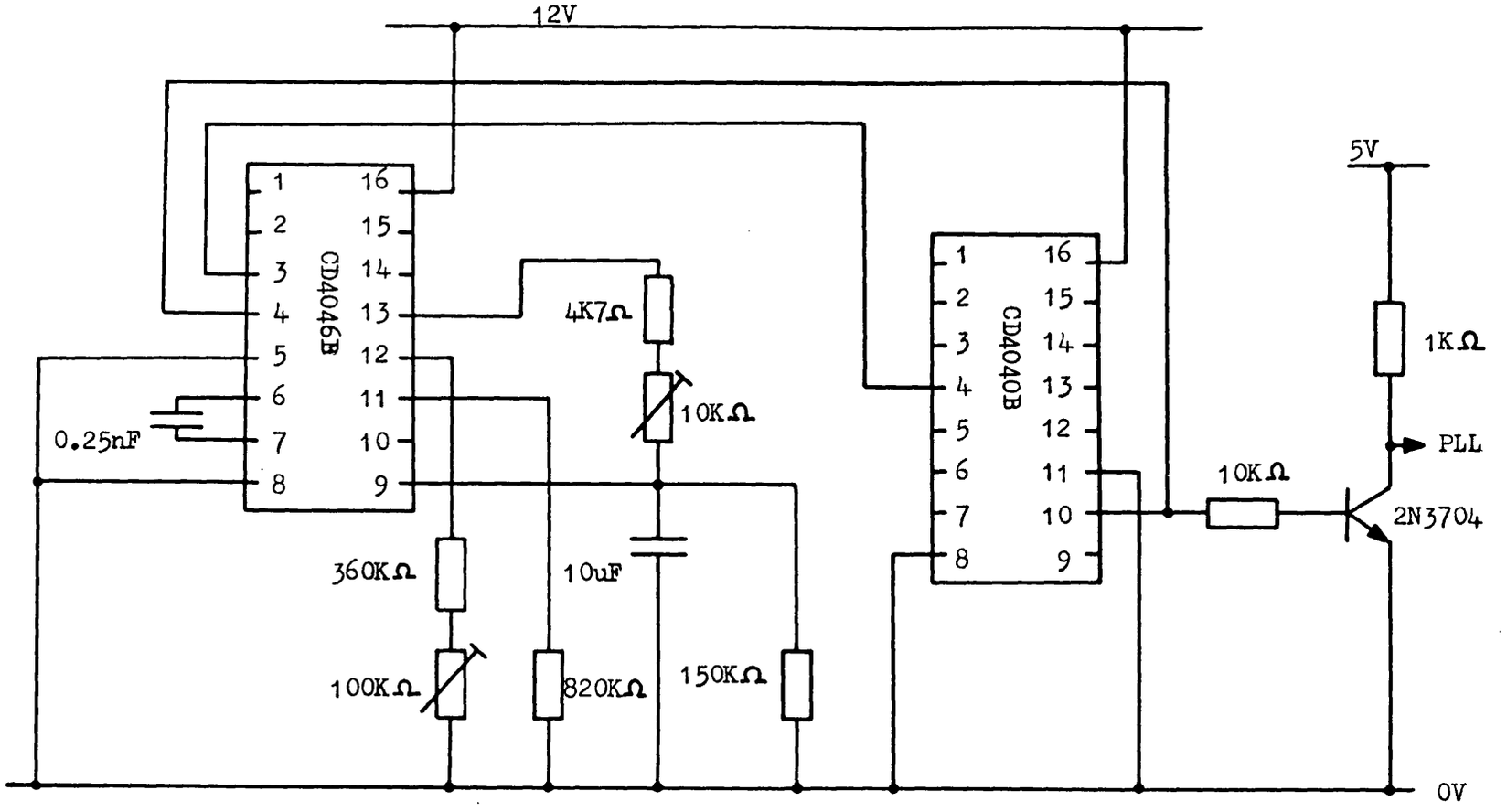


Fig. 1.2.3.c Digital counter and comparator circuit

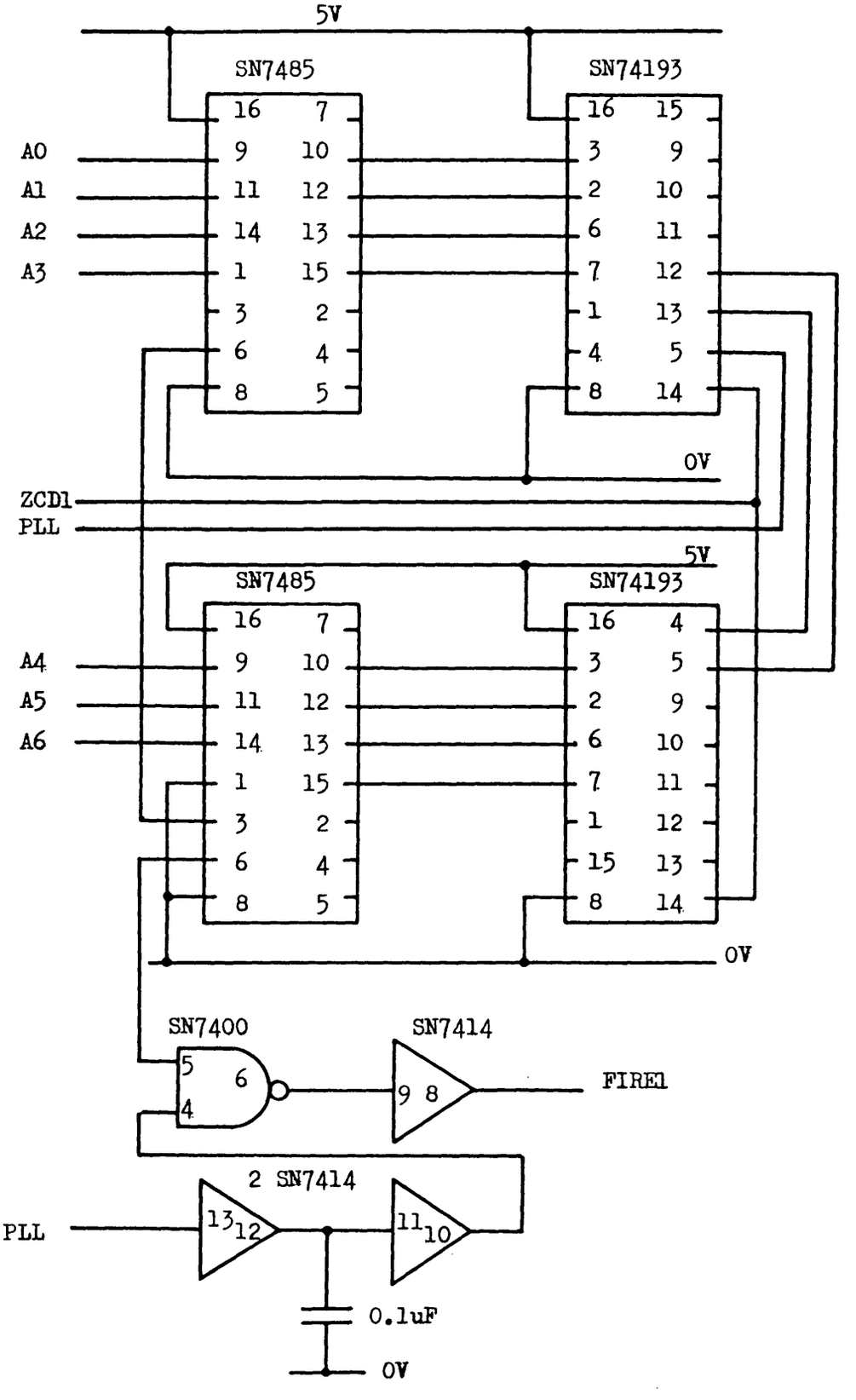


Fig. 1.2.3.d Triac firing and heater control circuit

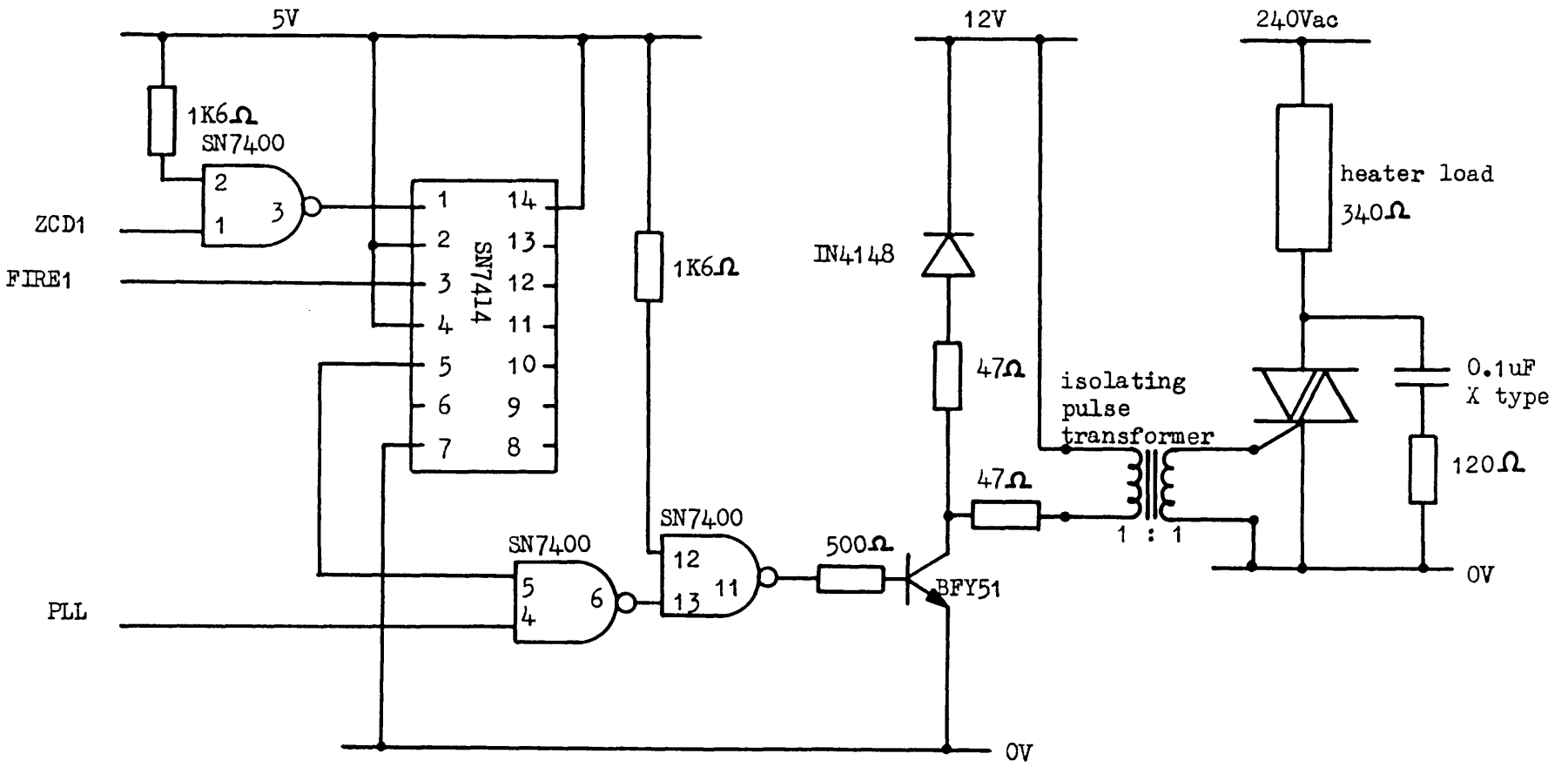


Fig. 1.2.3.e Disturbance lamp control circuit

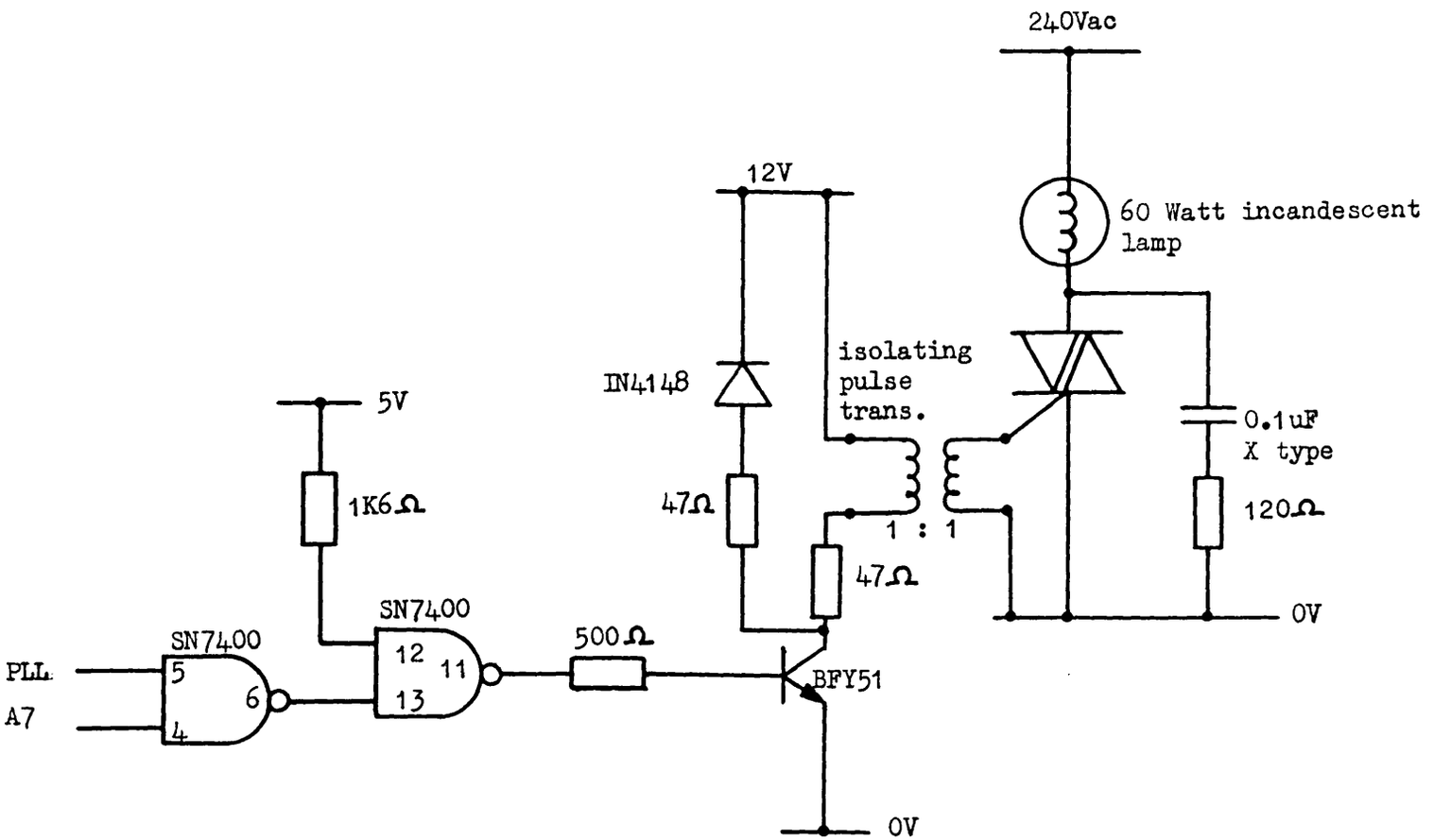


Fig. 1.2.4 Wheatstone bridge detector circuit

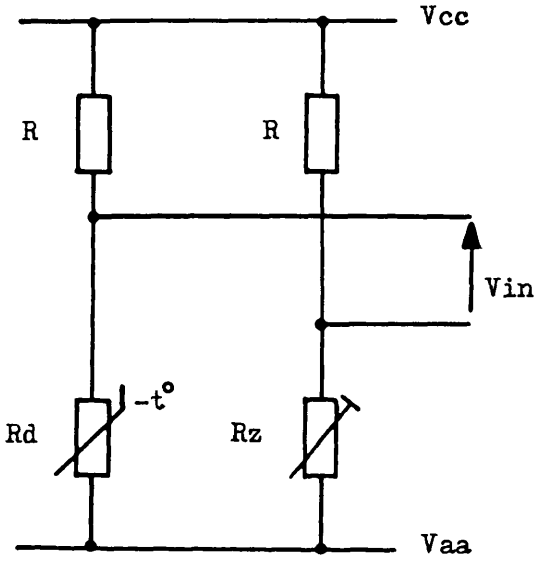


Fig. 1.2.5.a Detector bridge , multiplexor , and amplifier circuit

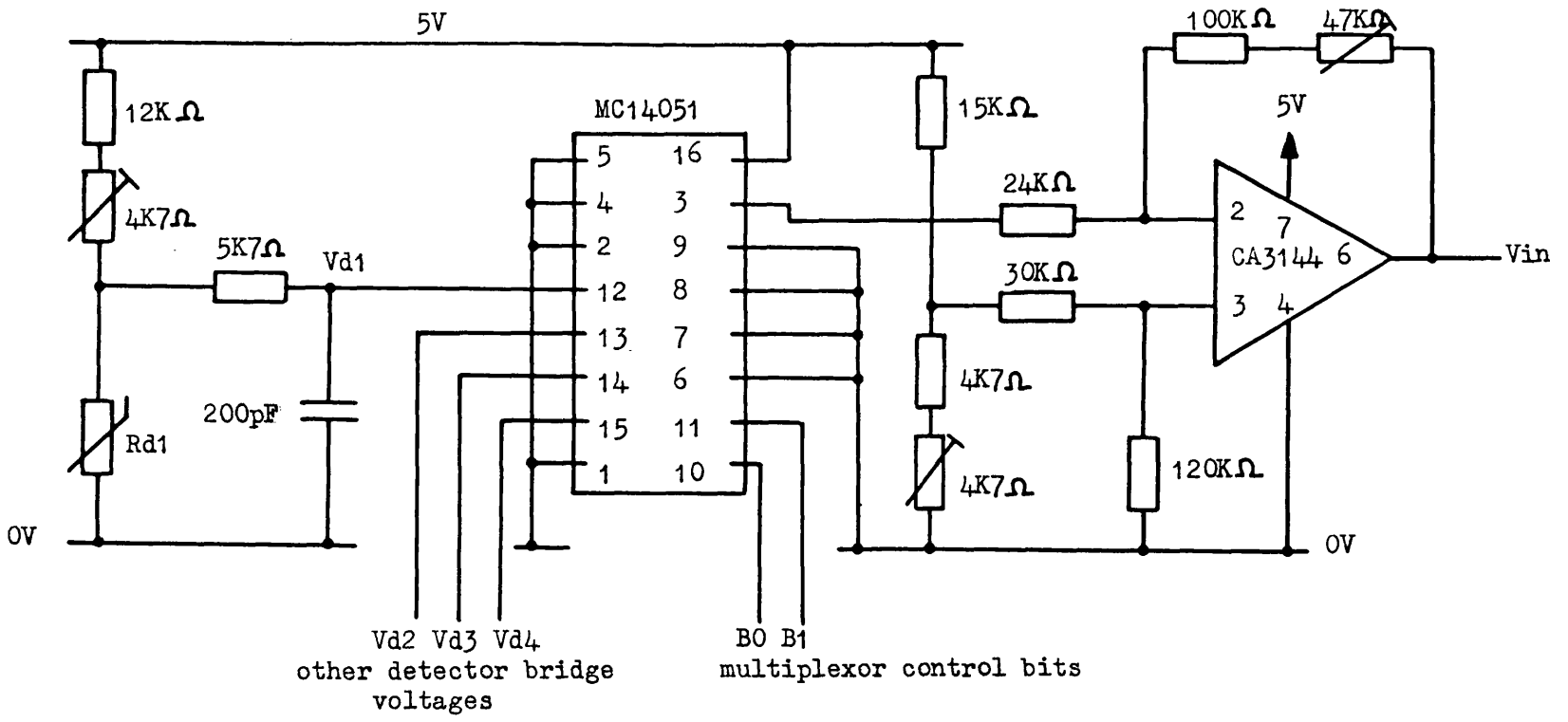


Fig. 1.2.5.b Analogue to digital converter

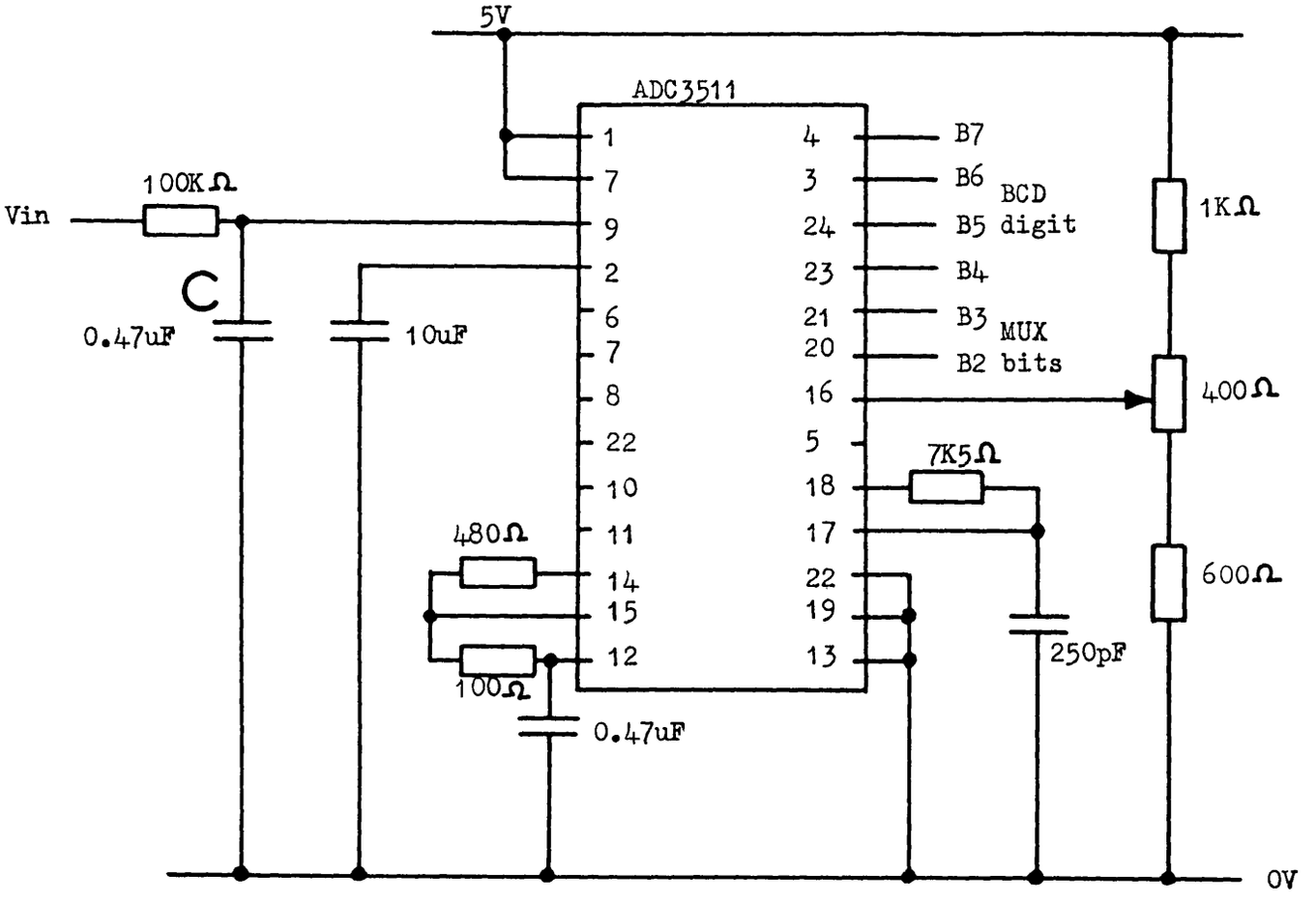
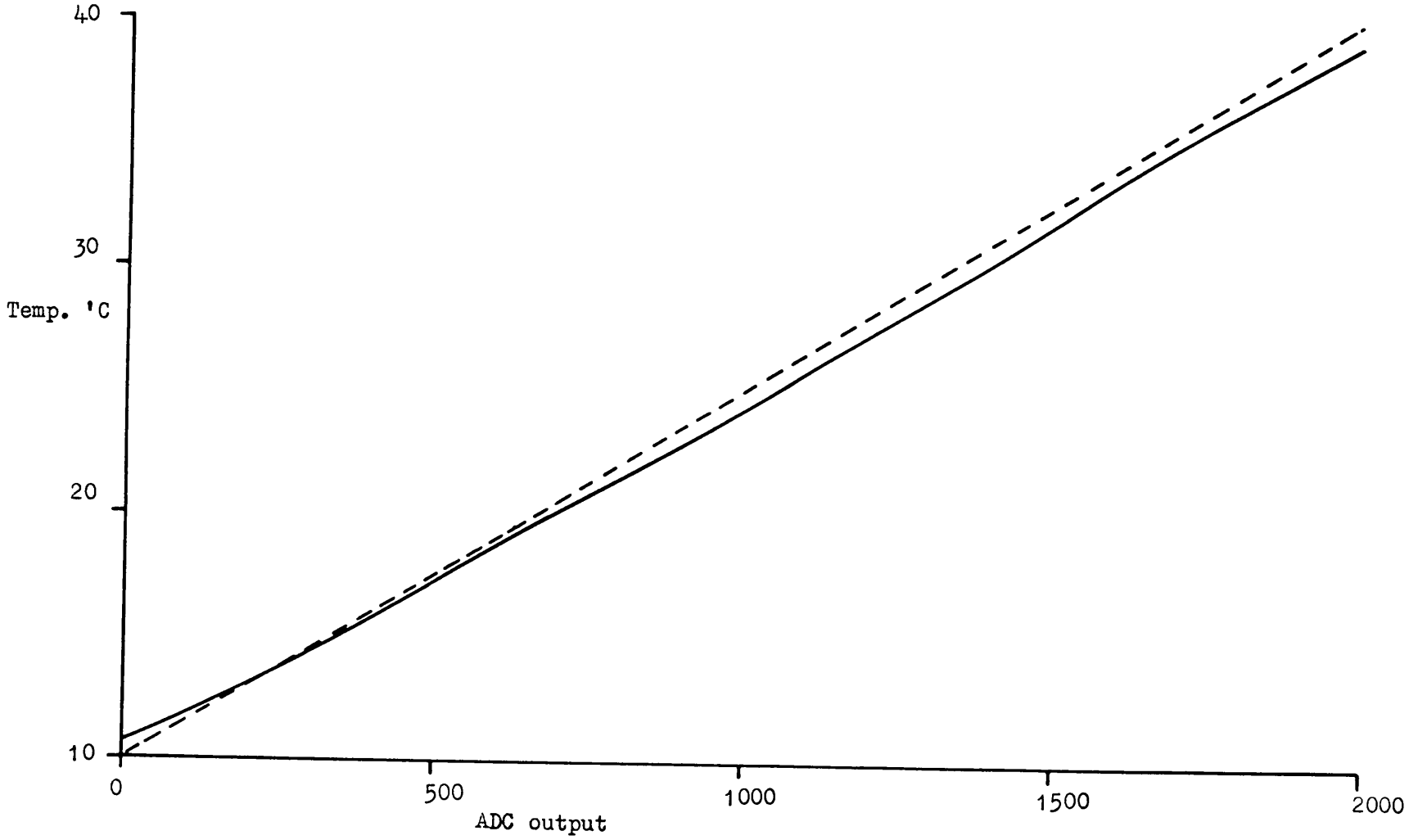


Fig. 1.2.6 ADC output against detector temperature



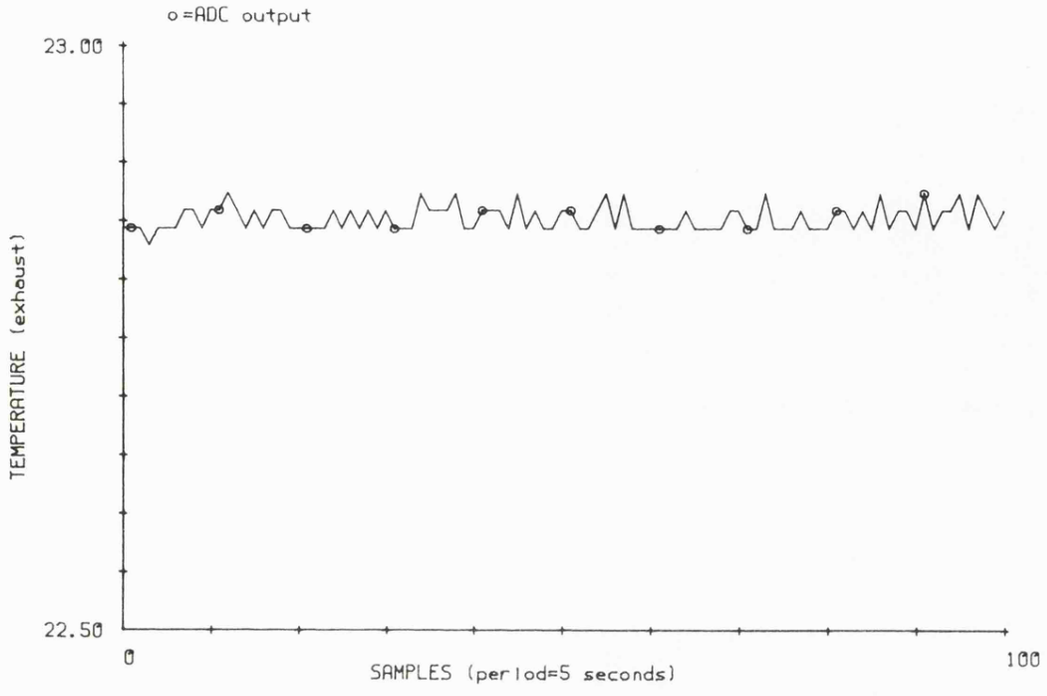


Fig. 1.2.7 ADC output for fixed detector resistance

Fig. 1.2.8 Step response for 2 different detector positions

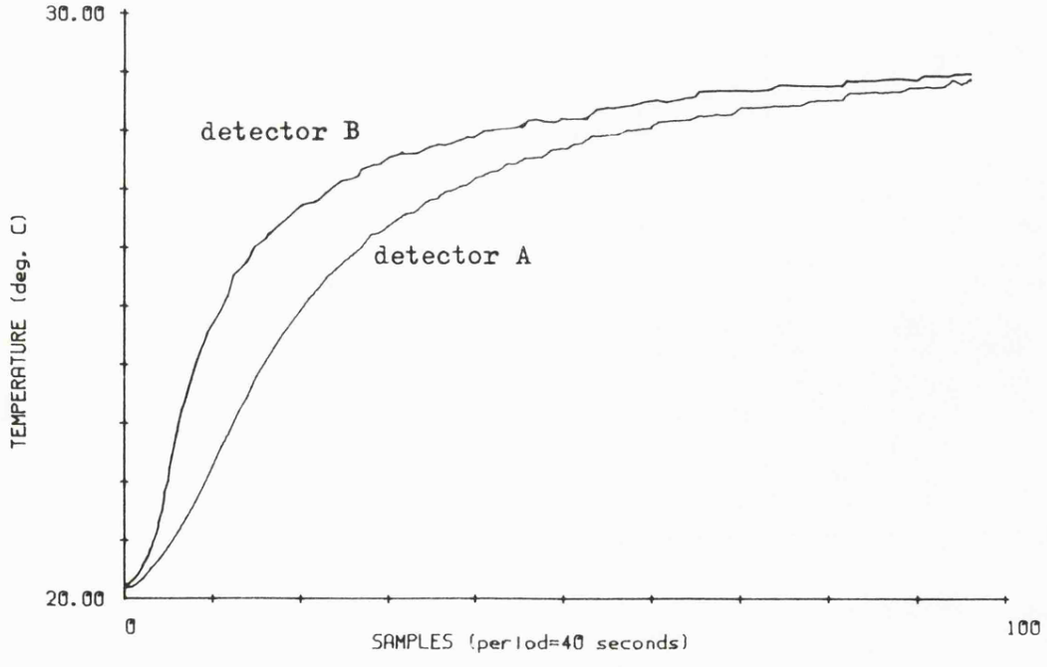


Fig. 2.1.1 Heat exchanger and 3 port mixing valve schematic diagram

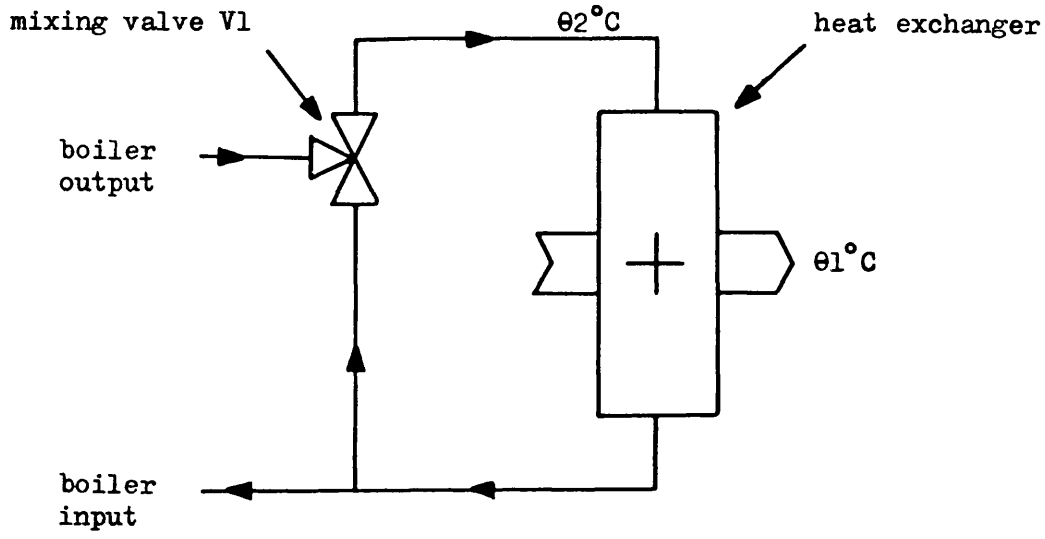


Fig. 2.1.2 Electrical analogy of air space thermodynamics

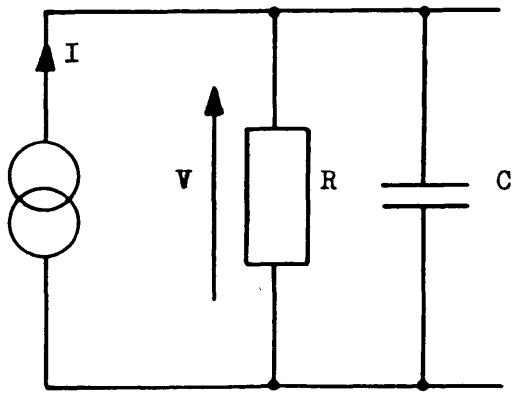


Fig. 2.1.3 Electrical anlogy of air space and building fabric thermodynamics

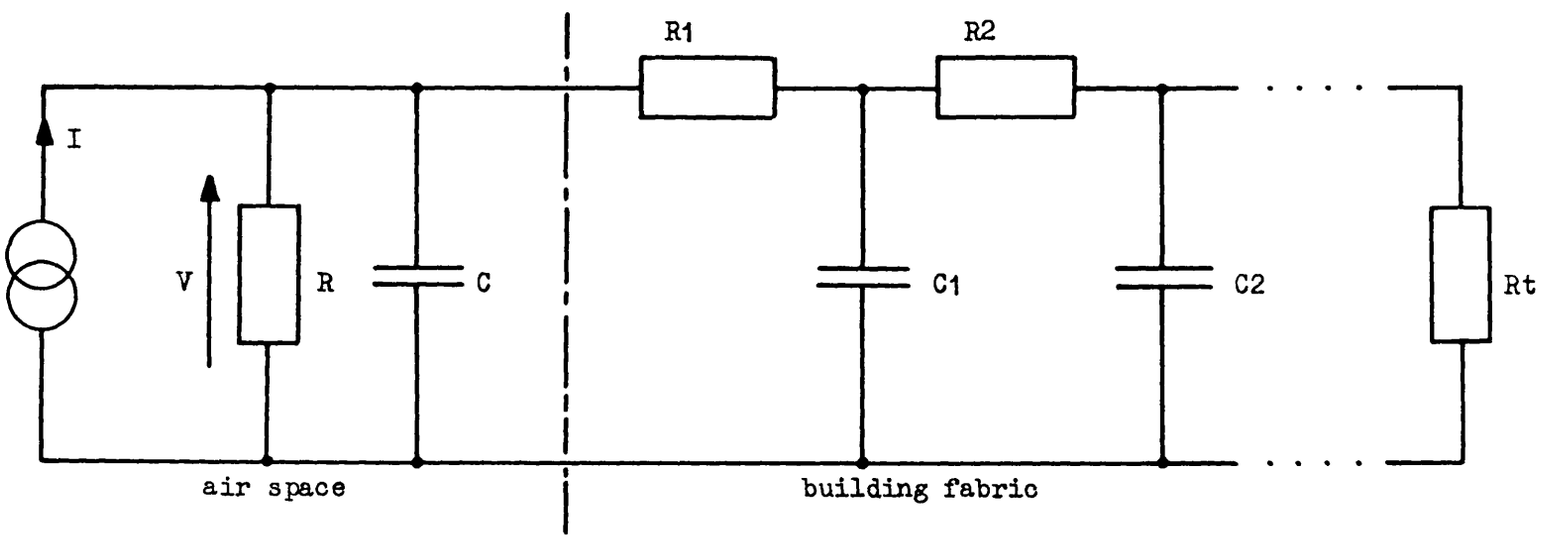
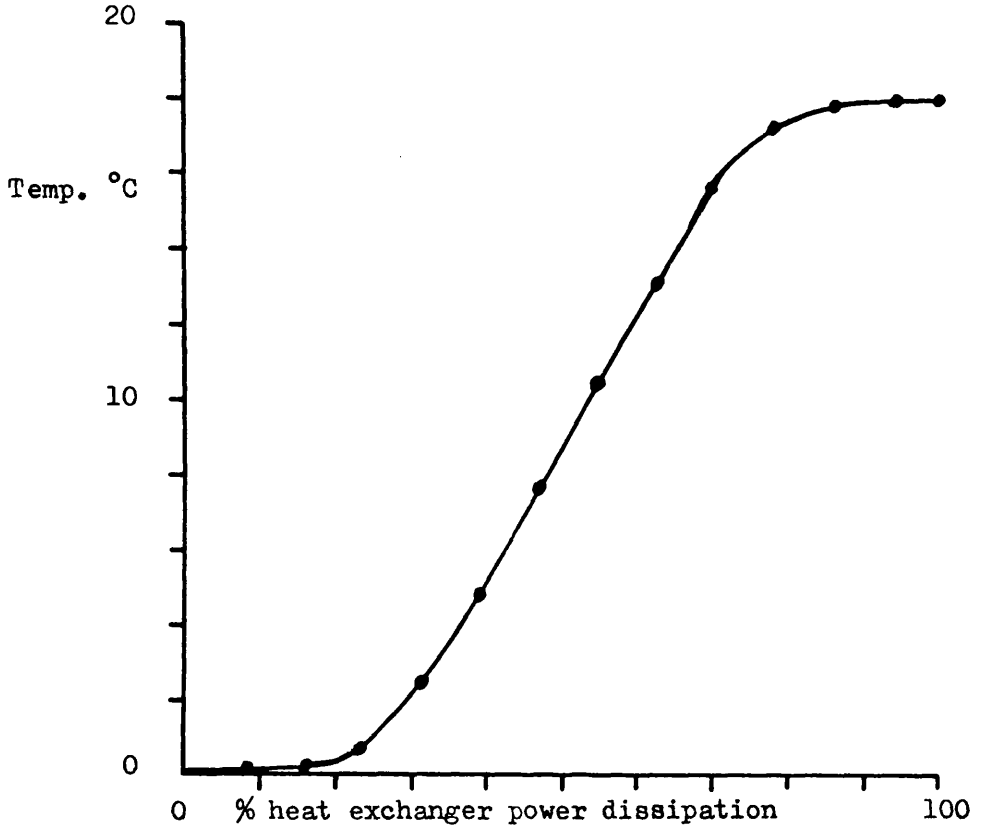


Fig. 2.2.1 Steady state increase in duct temperature for increasing power dissipated by heat exchanger



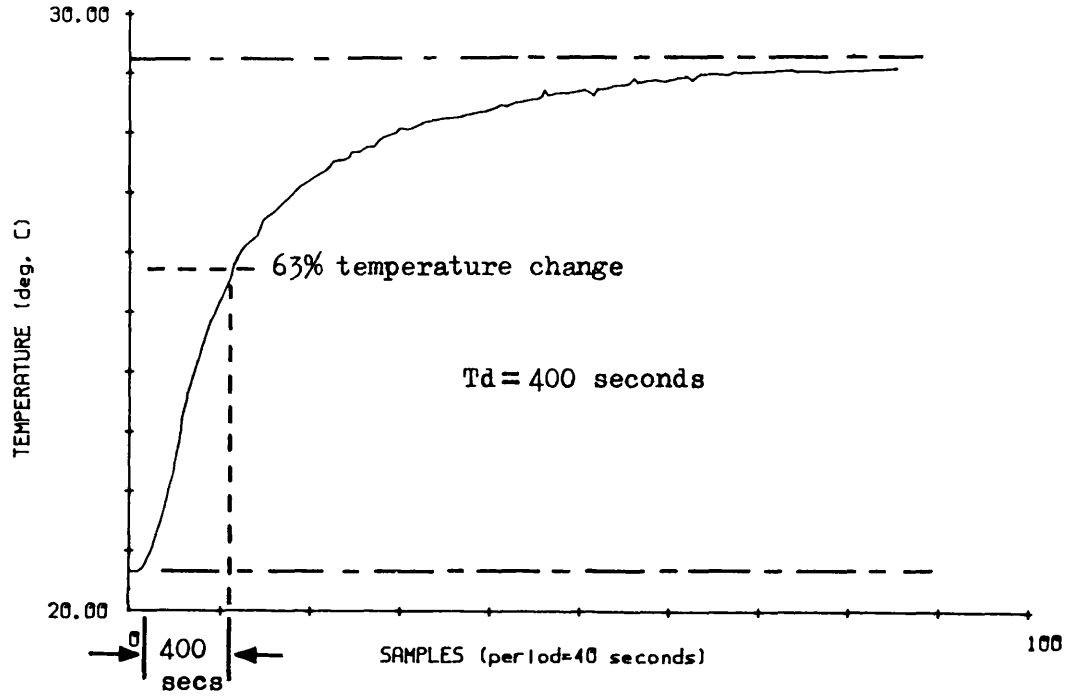


Fig. 2.2.2 Exhaust temperature step response

Fig. 2.2.3.a Weighting sequence from crosscorrelation of duct temperature

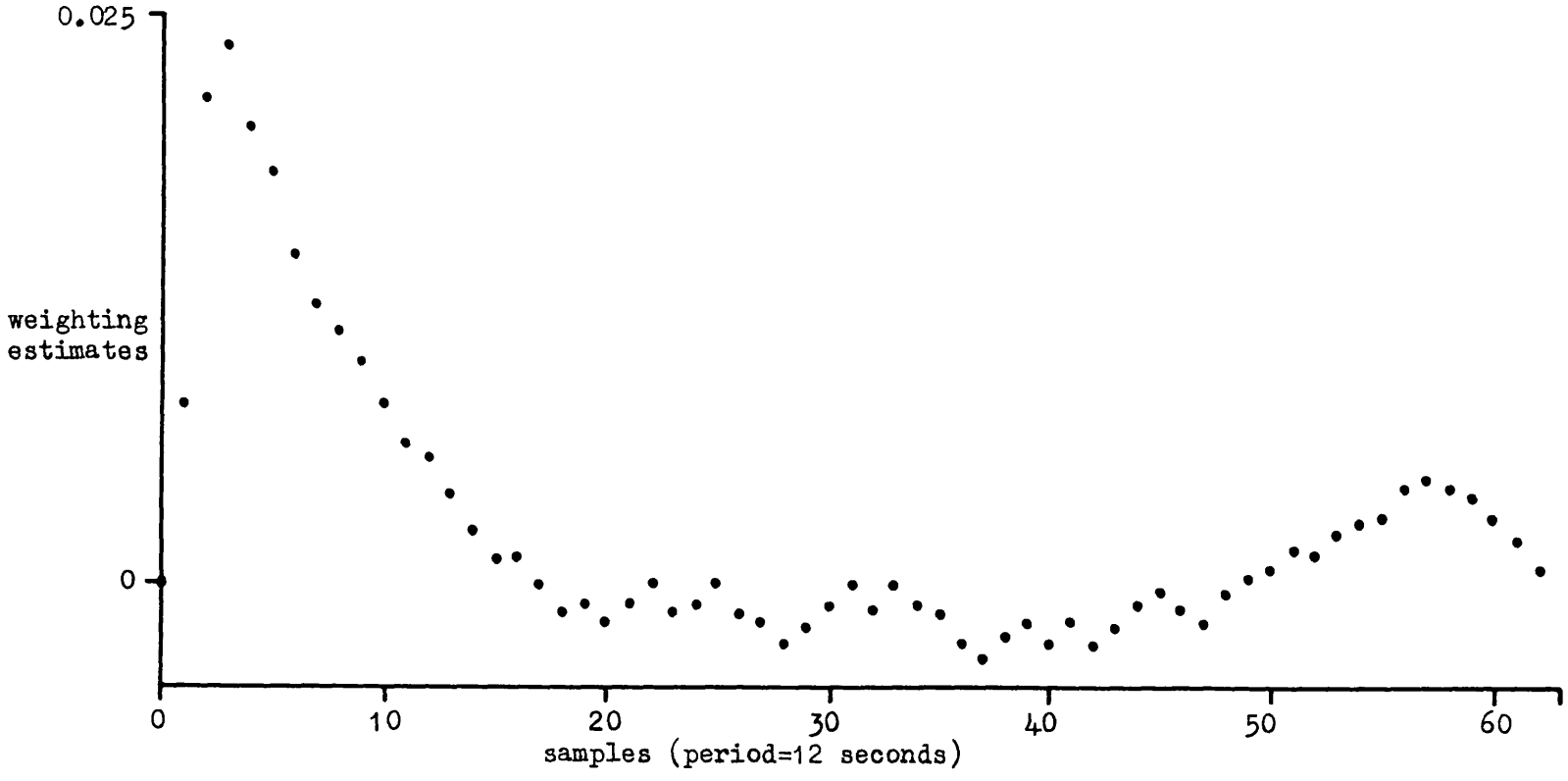


Fig. 2.2.3.b Weighting sequence from crosscorrelation of duct temperature using ambient temperature compensation

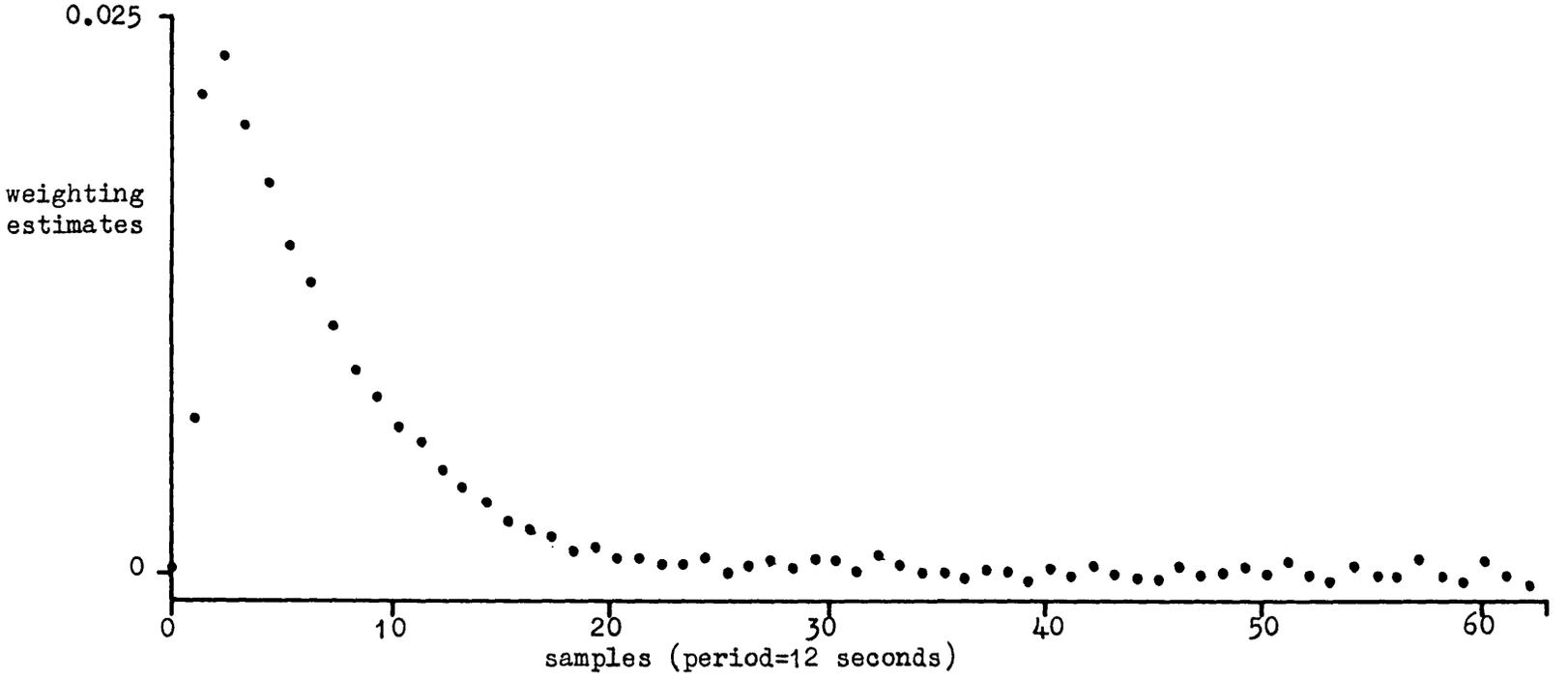


Fig. 2.2.4 Weighting sequence from crosscorrelation of exhaust temperature

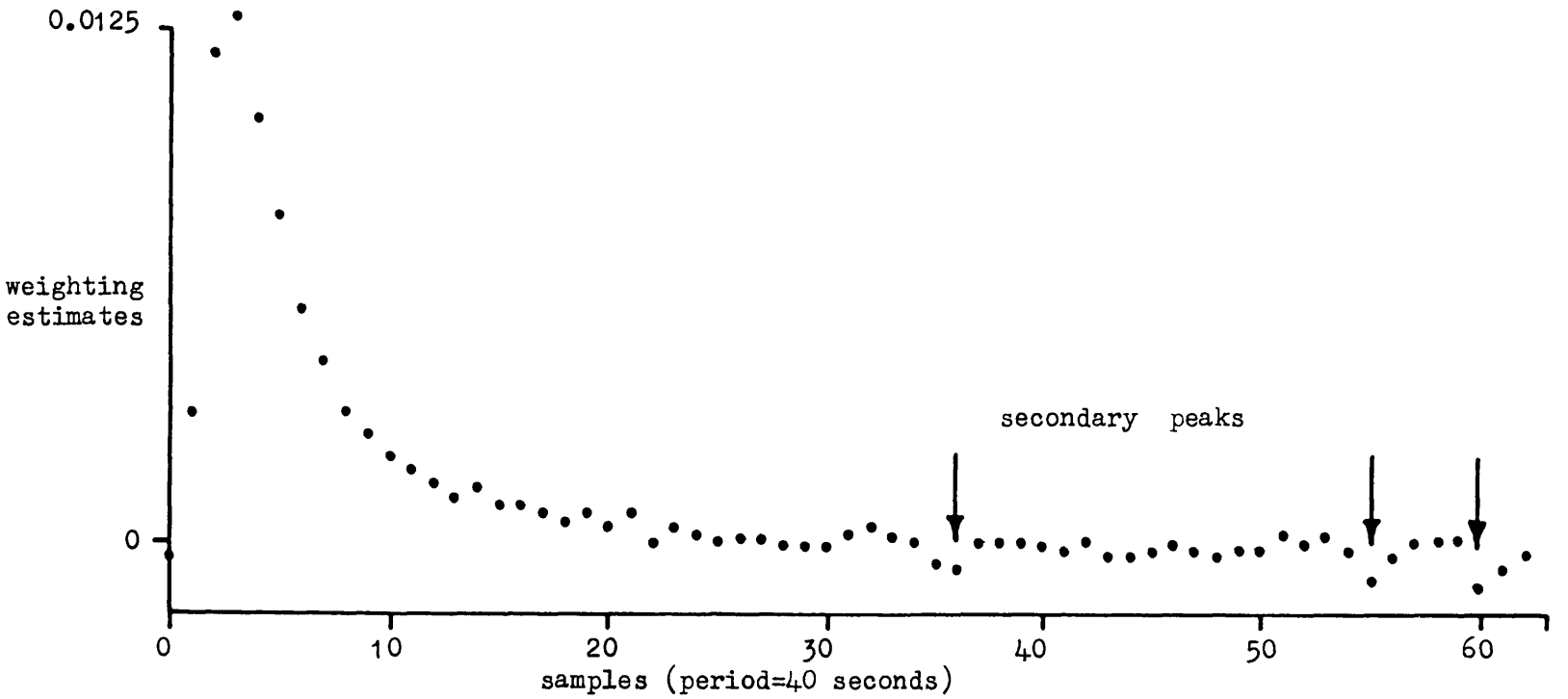


Fig. 2.2.5 Weighting sequence from crosscorrelation of exhaust temperature

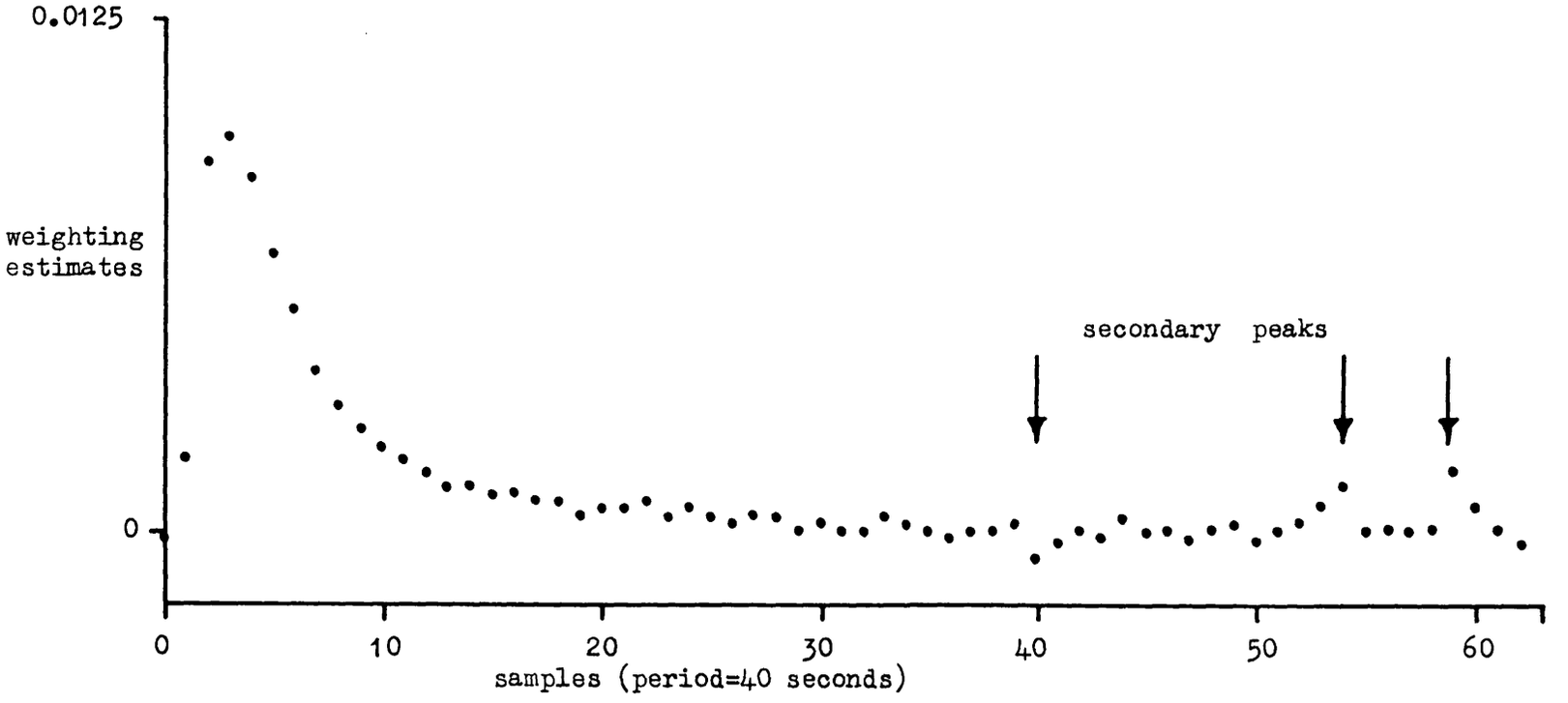
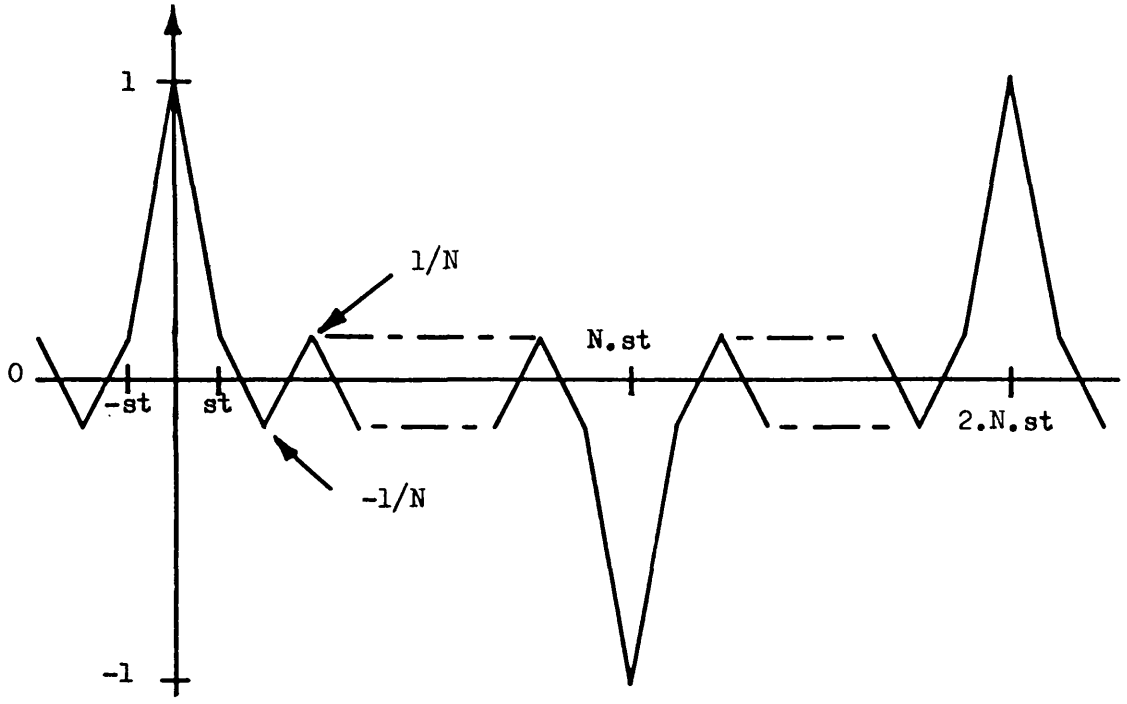


Fig. 2.2.6 Autocorrelation of Inverse Repeat PRBS



N is length of sequence
st is sample time

Fig. 2.2.7 Weighting sequence from crosscorrelation of exhaust temperature using Inverse Repeat PRBS

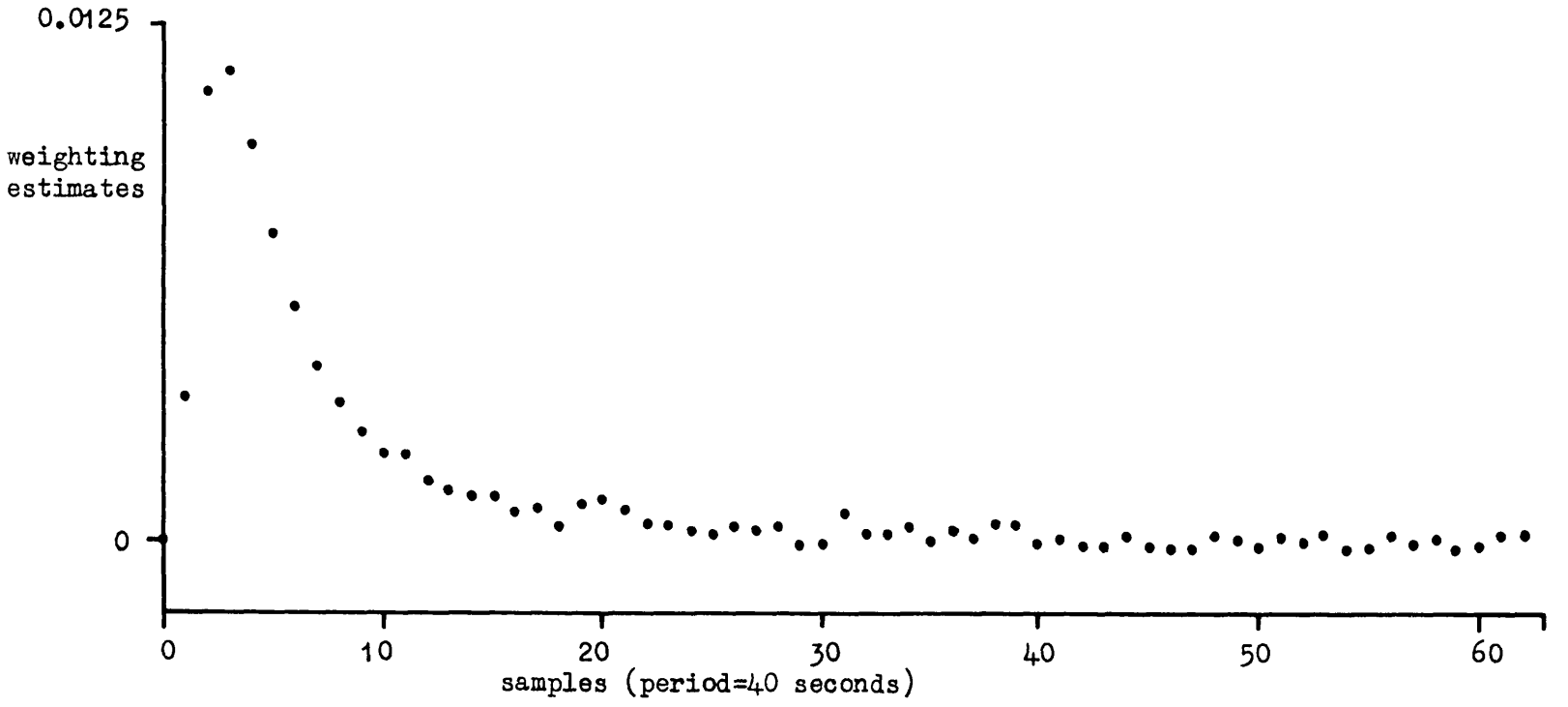


Fig. 2.2.8 Linear discrete model

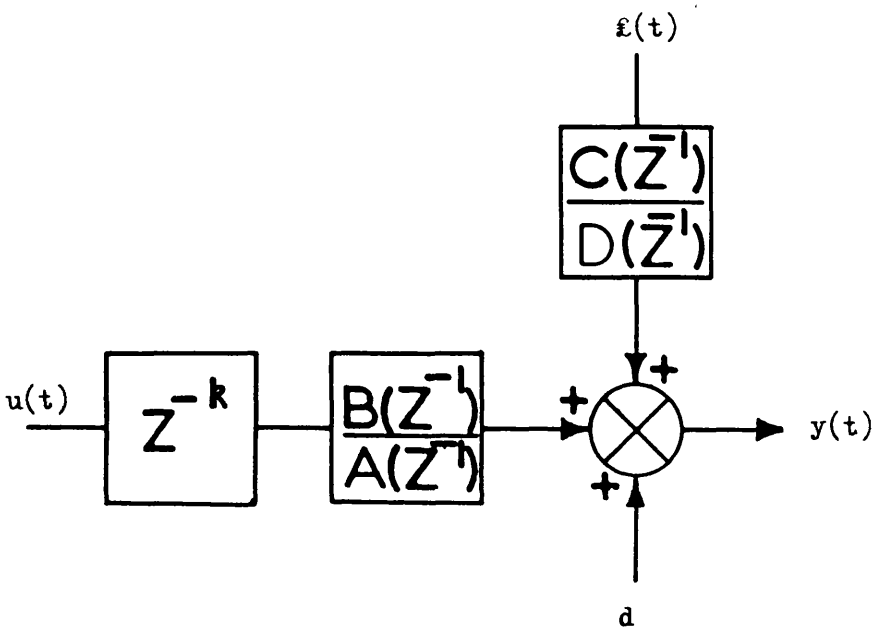


Fig. 2.3.1.a Pulse response from weighting sequence and identified transfer function

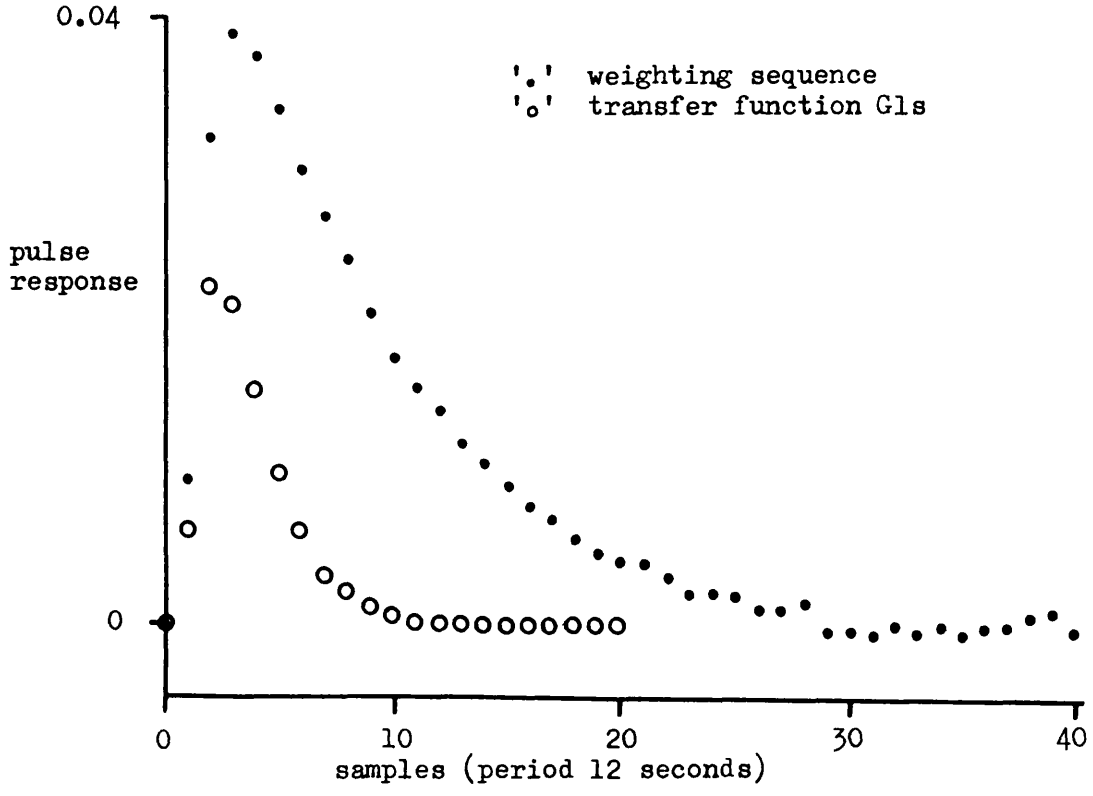


Fig. 2.3.1.b Pulse response from weighting sequence and identified transfer function

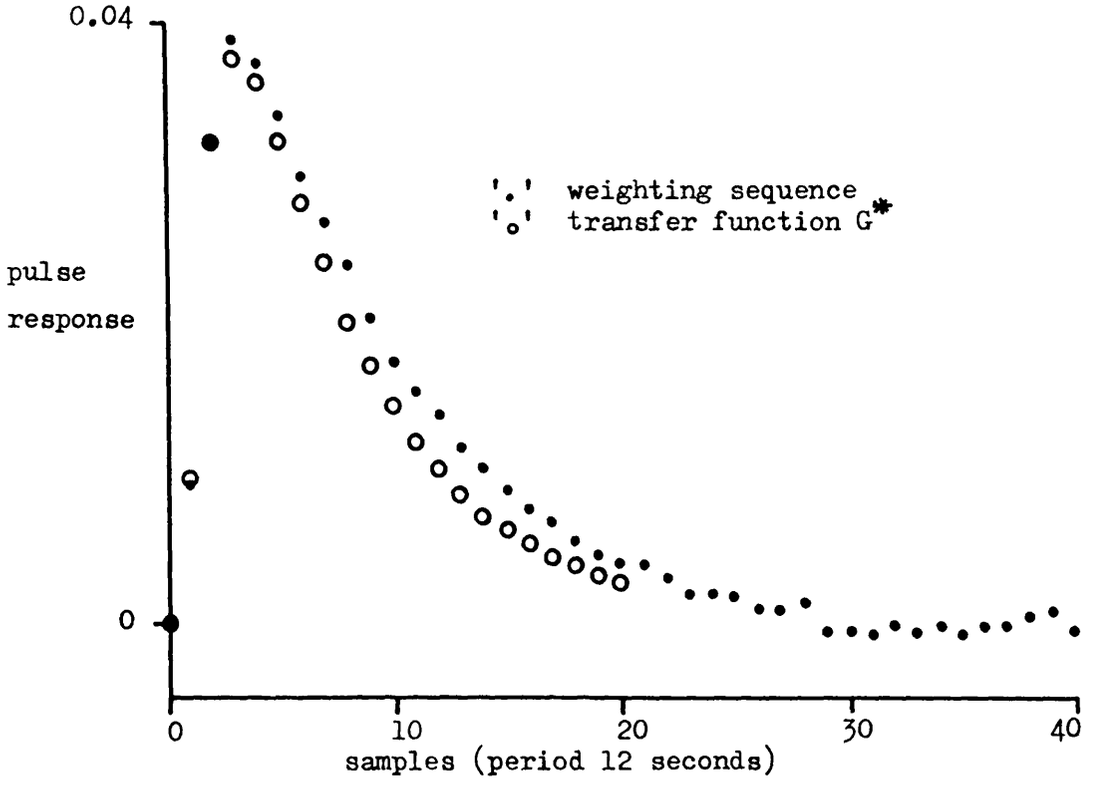


Fig. 2.3.1.c

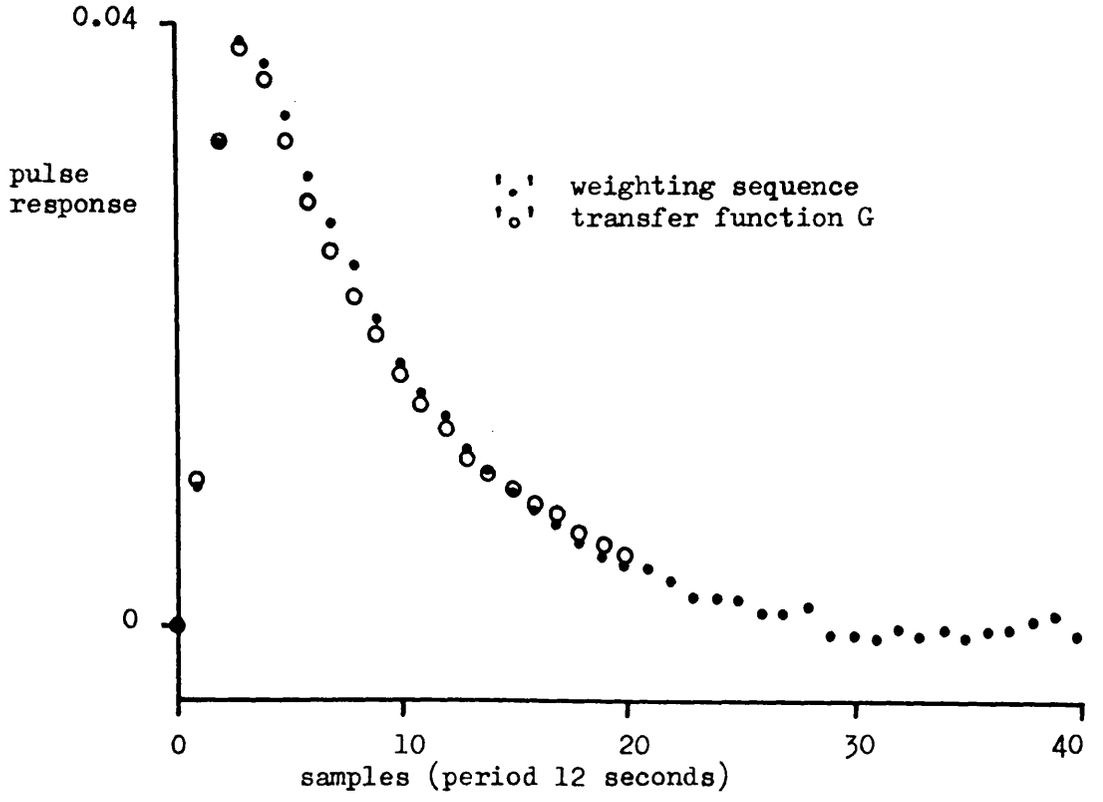


Fig. 2.3.2.a Pulse response from weighting sequence and identified transfer function

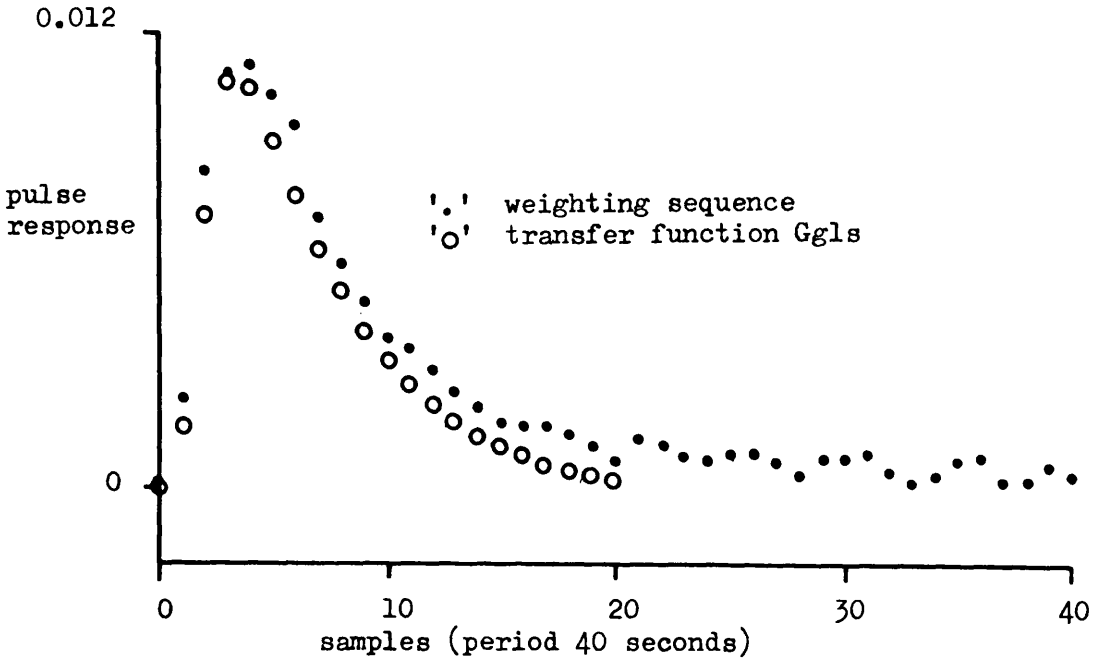


Fig. 2.3.2.b Pulse response from weighting sequence and identified transfer function

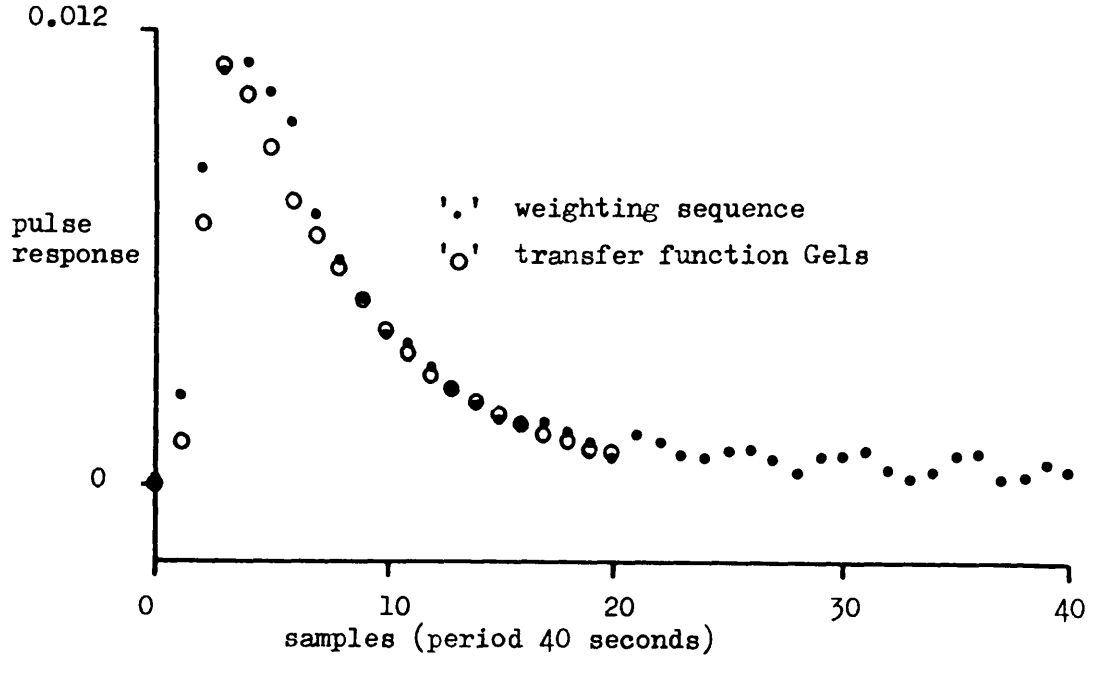


Fig. 2.3.3.a Pulse response from weighting sequence and identified transfer function

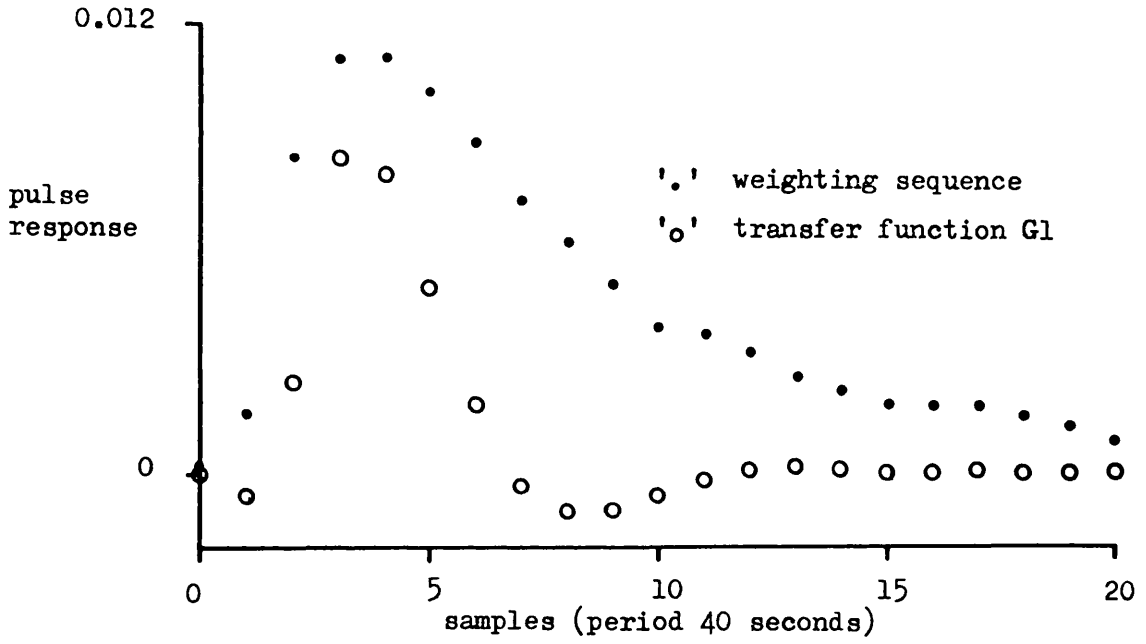


Fig. 2.3.3.b Pulse response from weighting sequence and identified transfer function

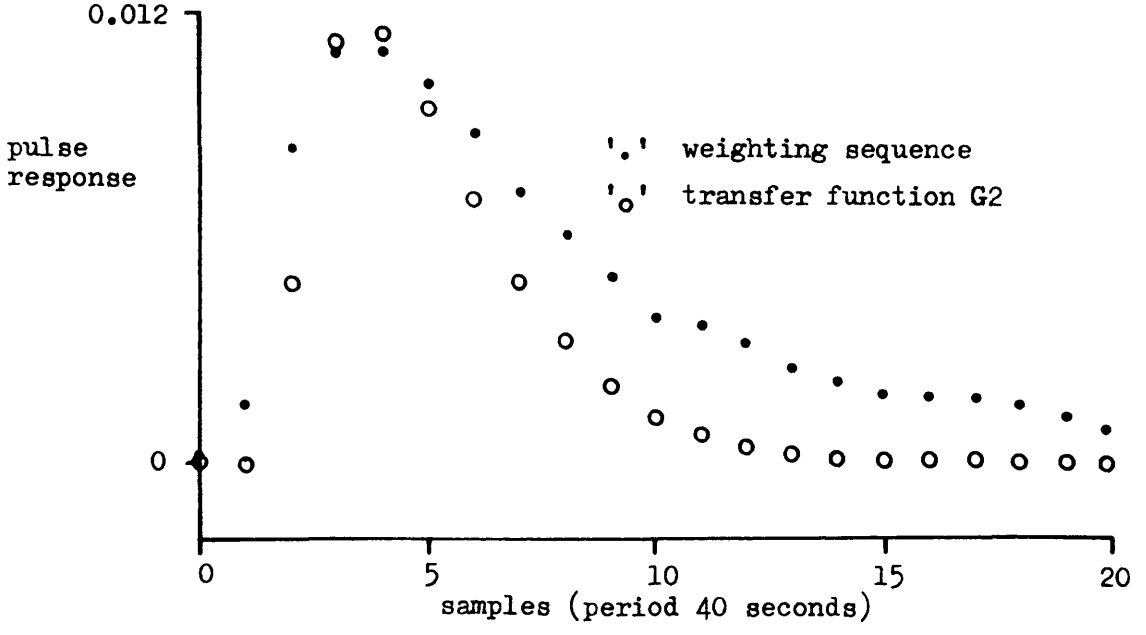


Fig. 2.3.3.c Pulse response from weighting sequence and identified transfer function

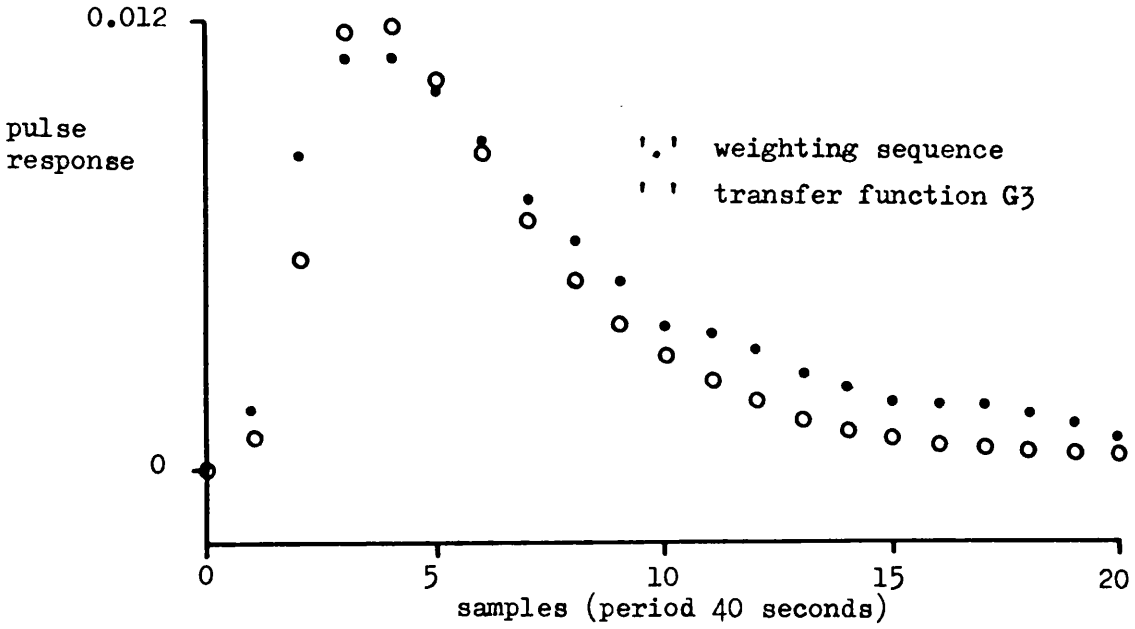


Fig. 2.3.4.a Pulse response from weighting sequence and identified transfer function

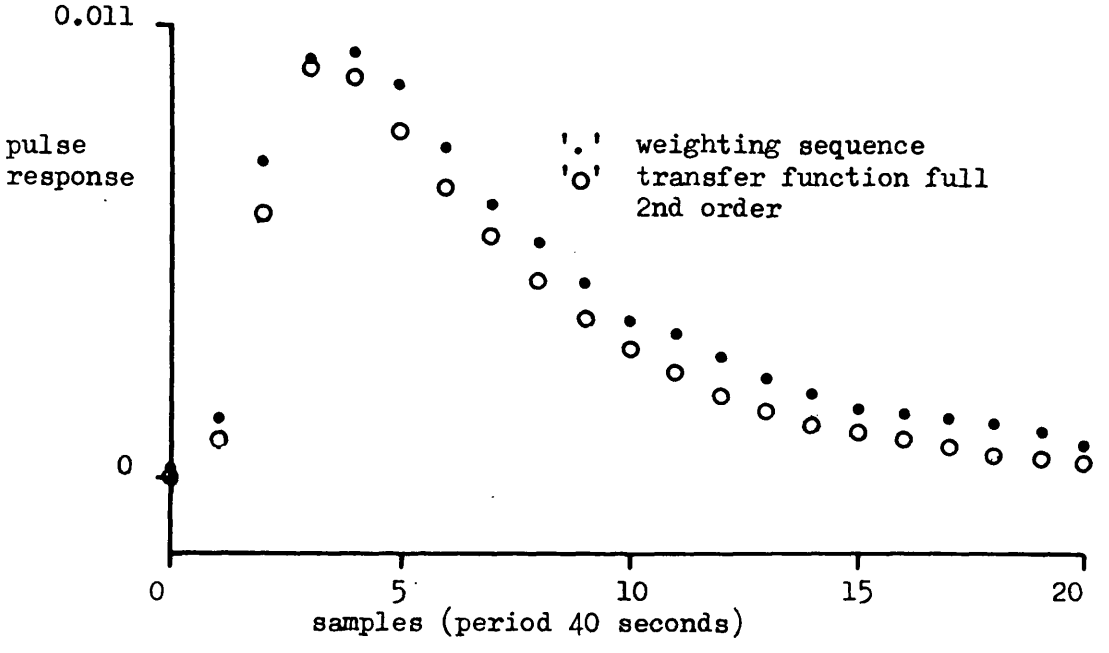


Fig. 2.3.4.b Pulse response from weighting sequence and identified transfer function

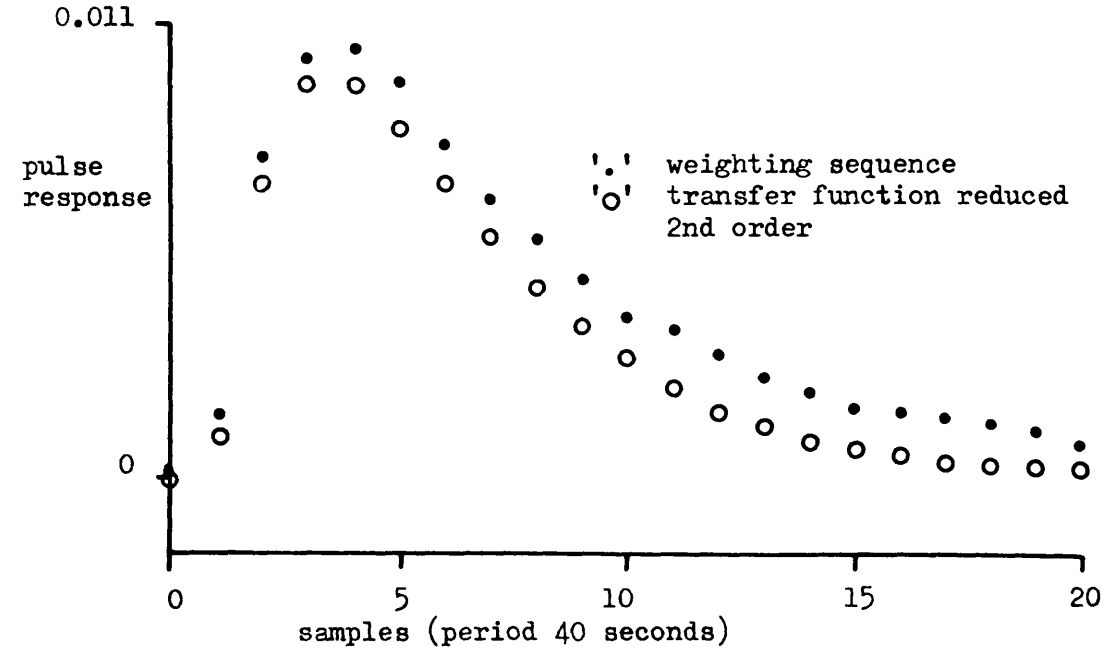


Fig. 2.3.4.c Pulse response from weighting sequence and simple first order model

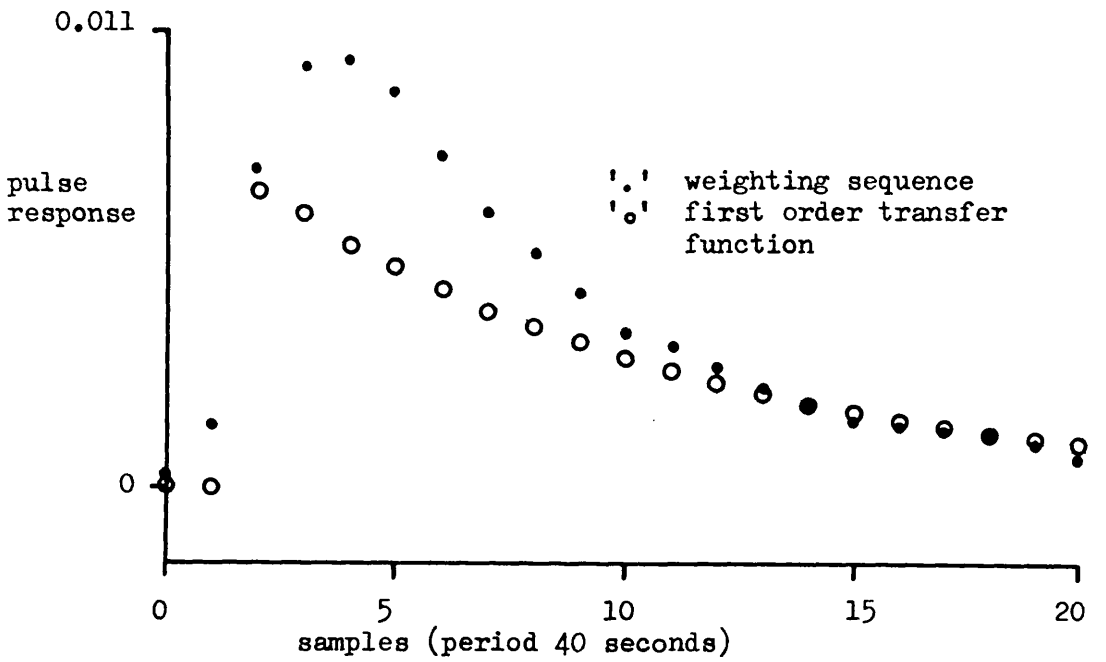


Fig. 2.3.5 Closed loop control system

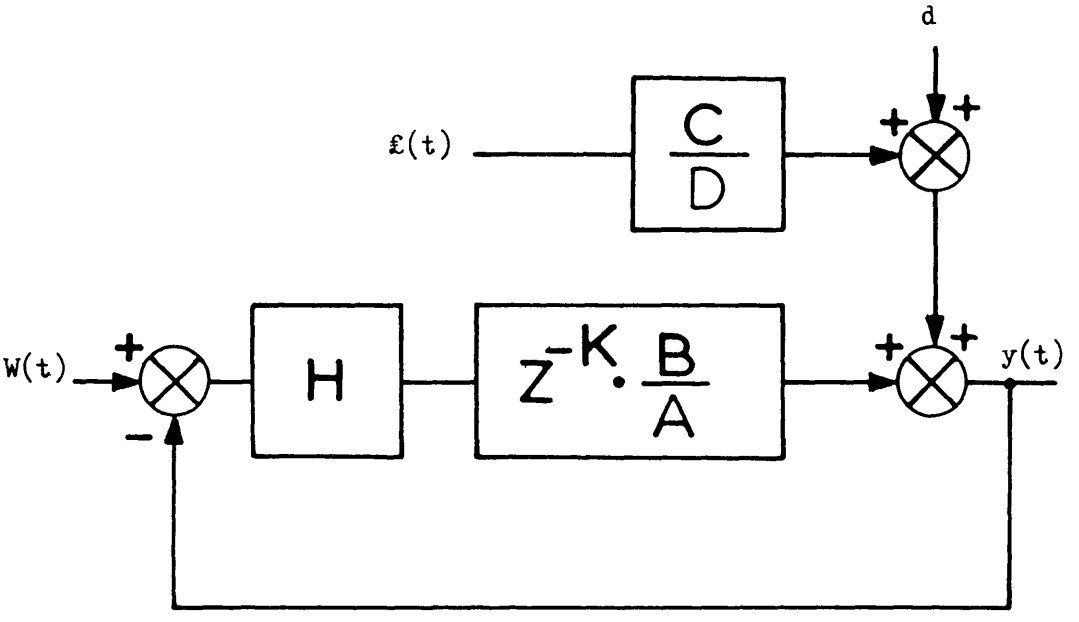


Fig. 2.3.6 Closed loop regulation system

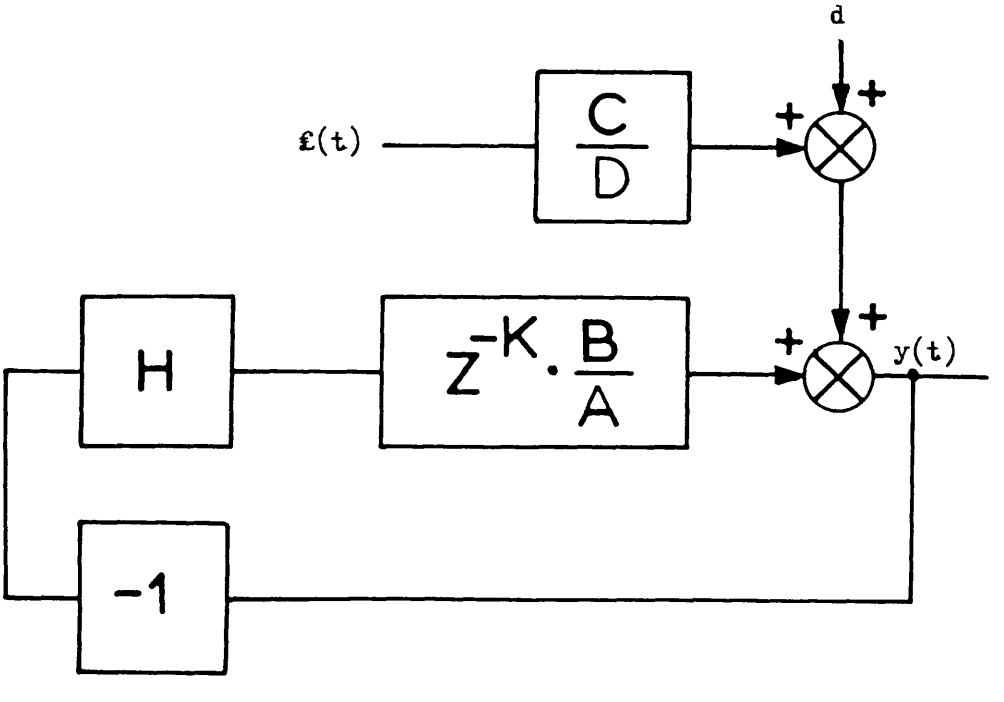


Fig. 3.1.1 Block diagram of time delay plus time lag plant under proportional control

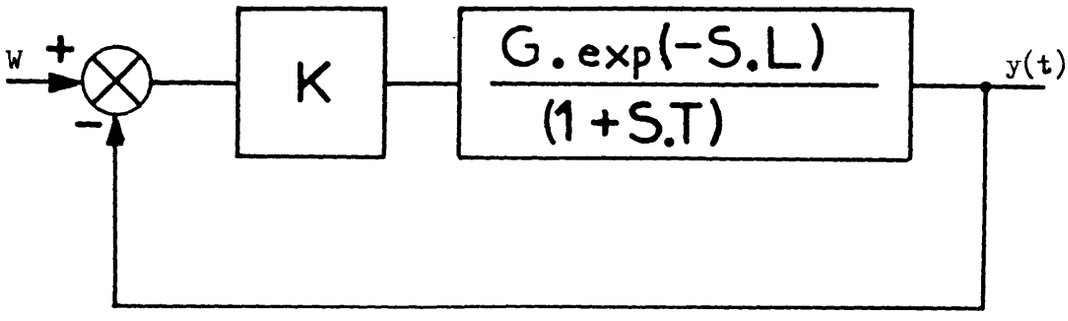
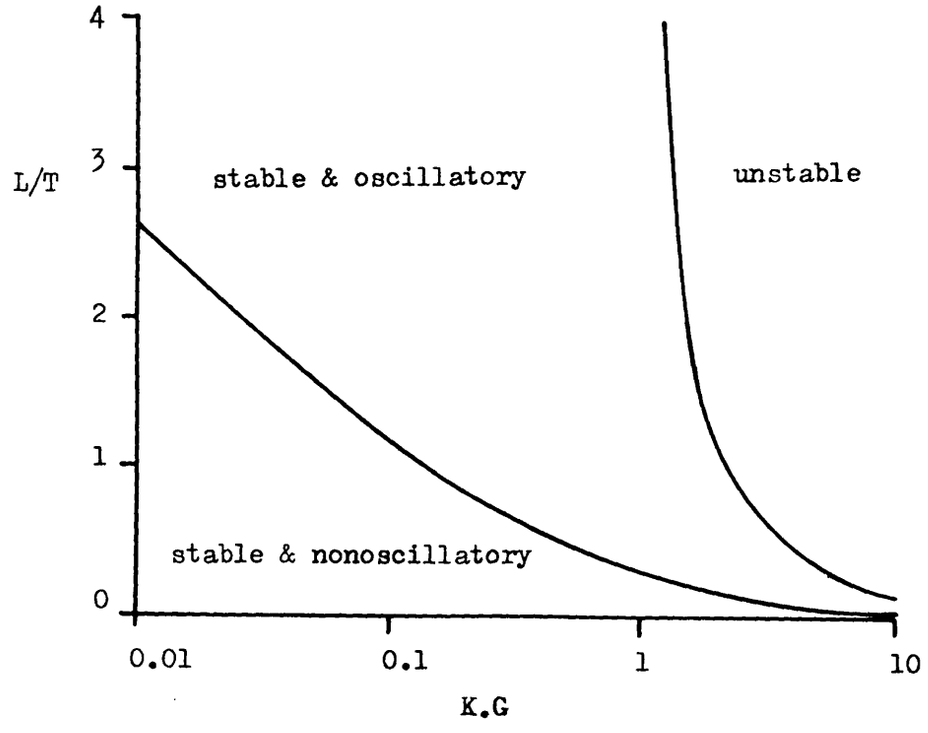


Fig. 3.1.2 Stability boundary with respect to open loop gain and controllability ratio



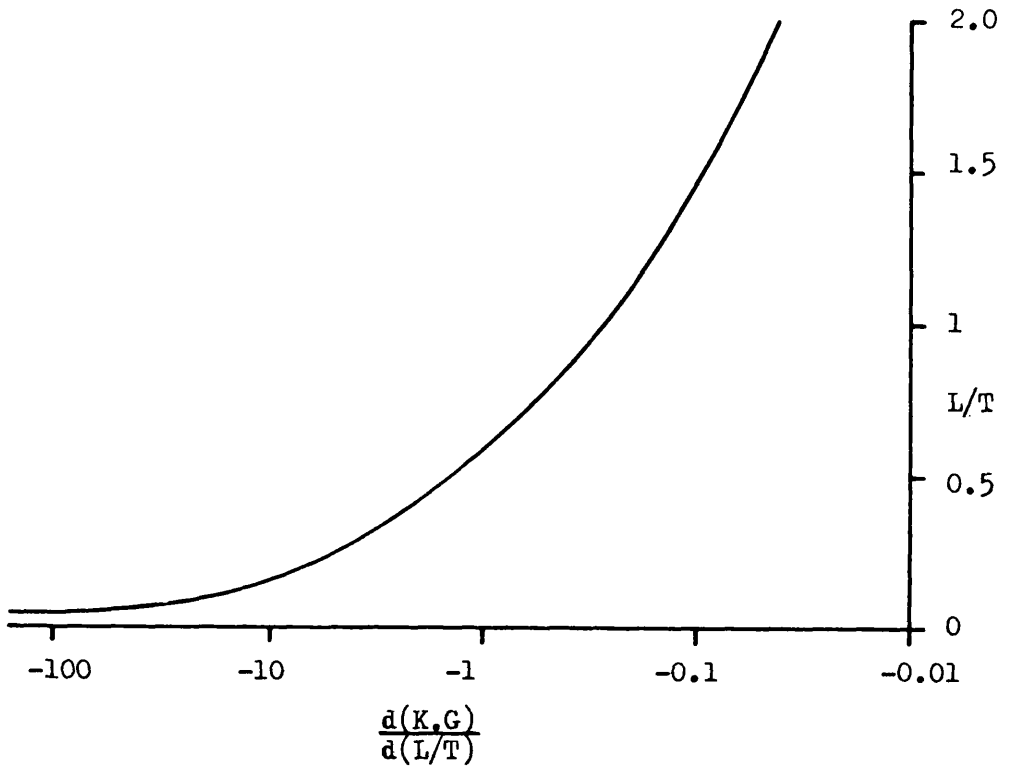


Fig. 3.1.3 Locus of controllability ratio with respect to gain sensitivity for a critically damped response

Fig. 3.2.1 Two equivalent control loops

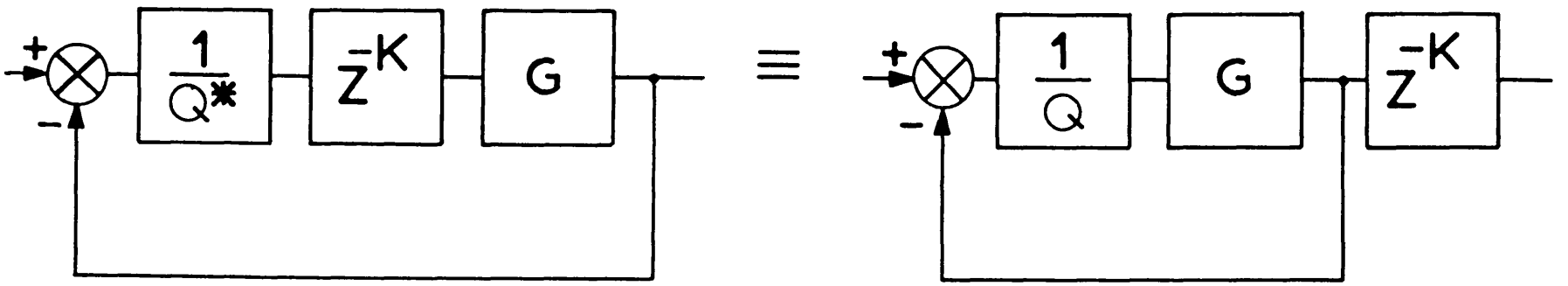


Fig. 3.2.2 Smith predictor control loop

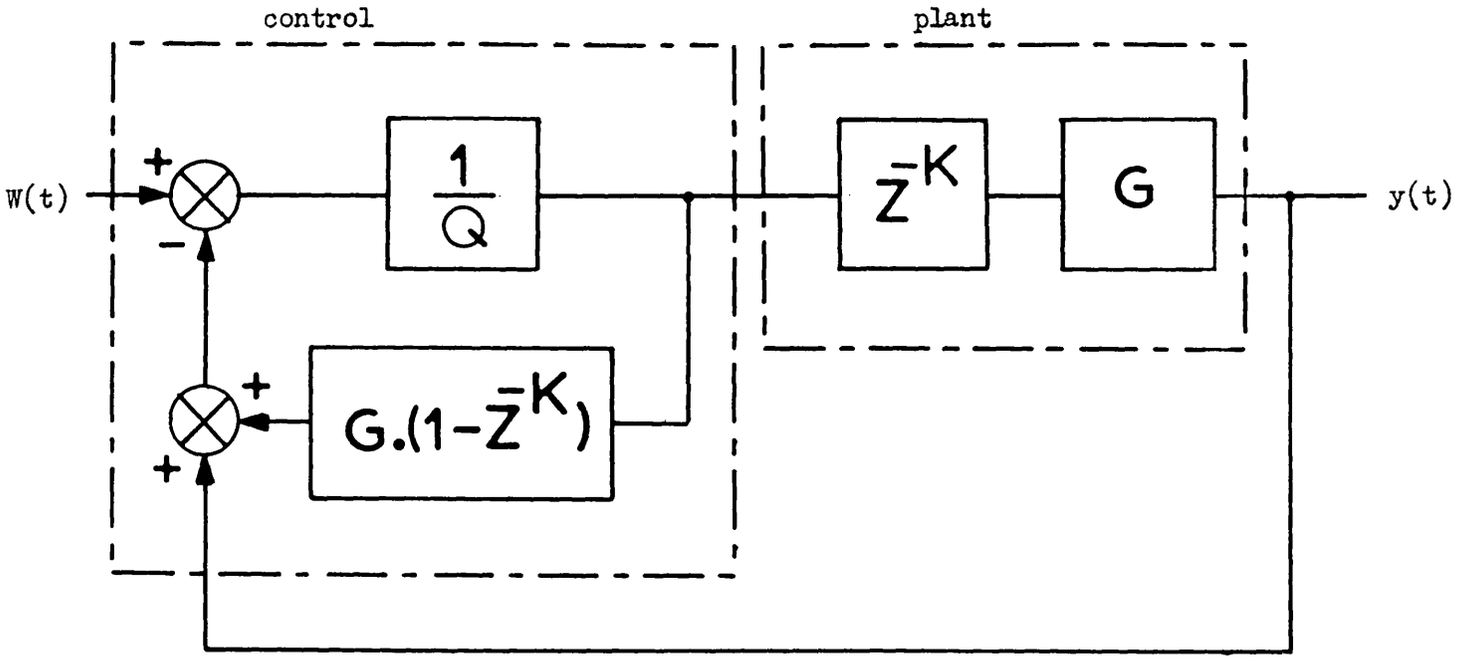


Fig 3.2.3 Least Squares predictor control loop

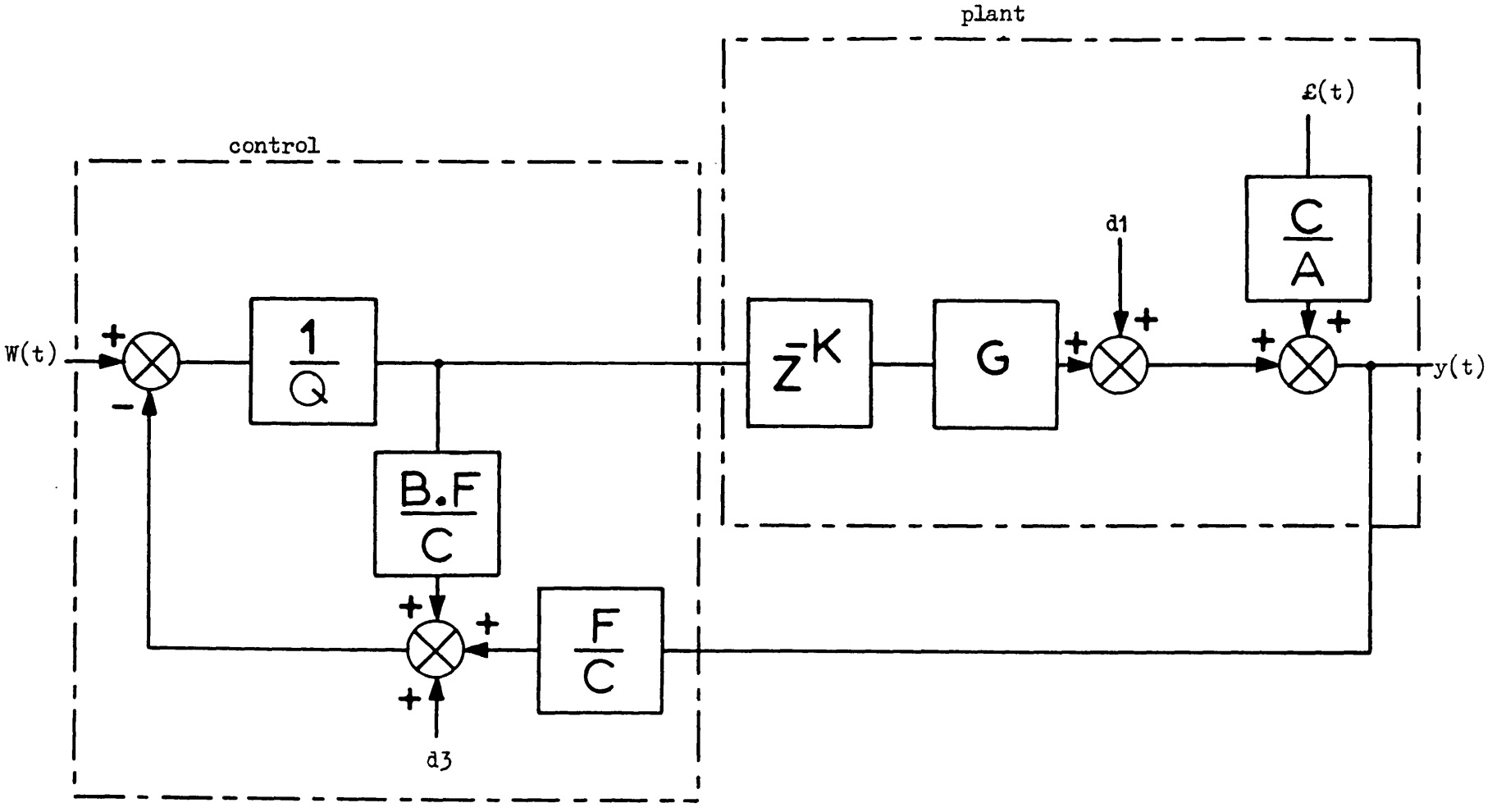


Fig. 4.1 Structure of the self-tuning controller

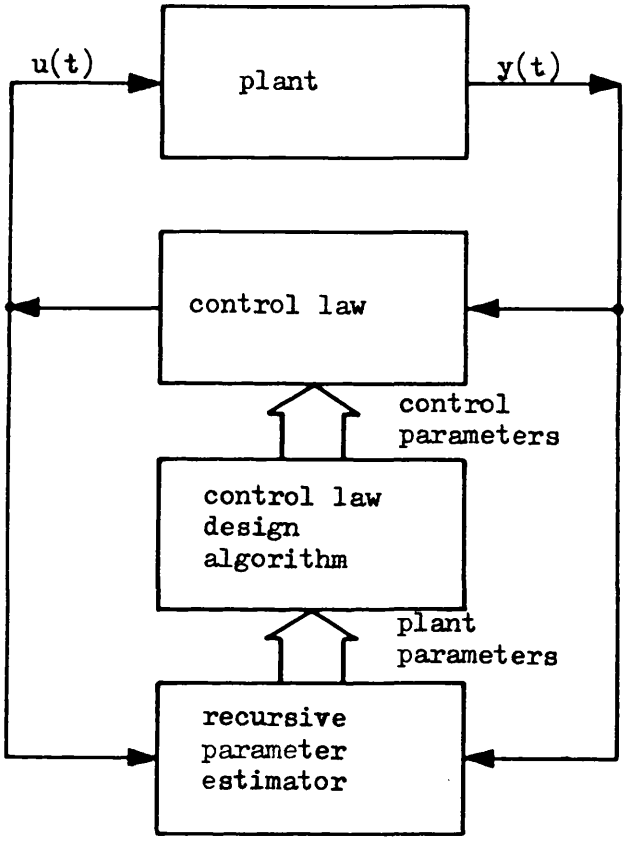
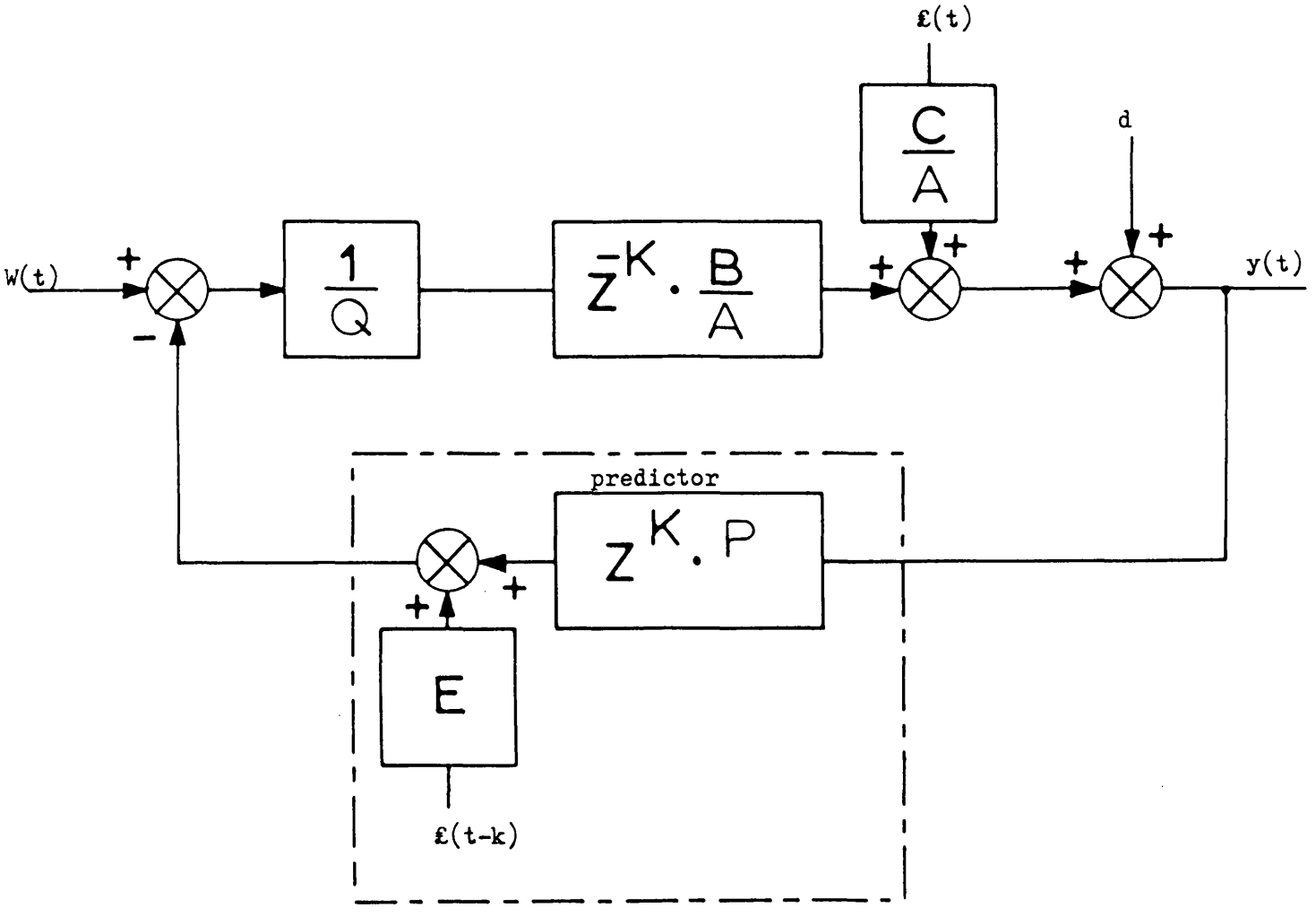


Fig. 4.1.1 Block diagram of least-squares optimal control loop



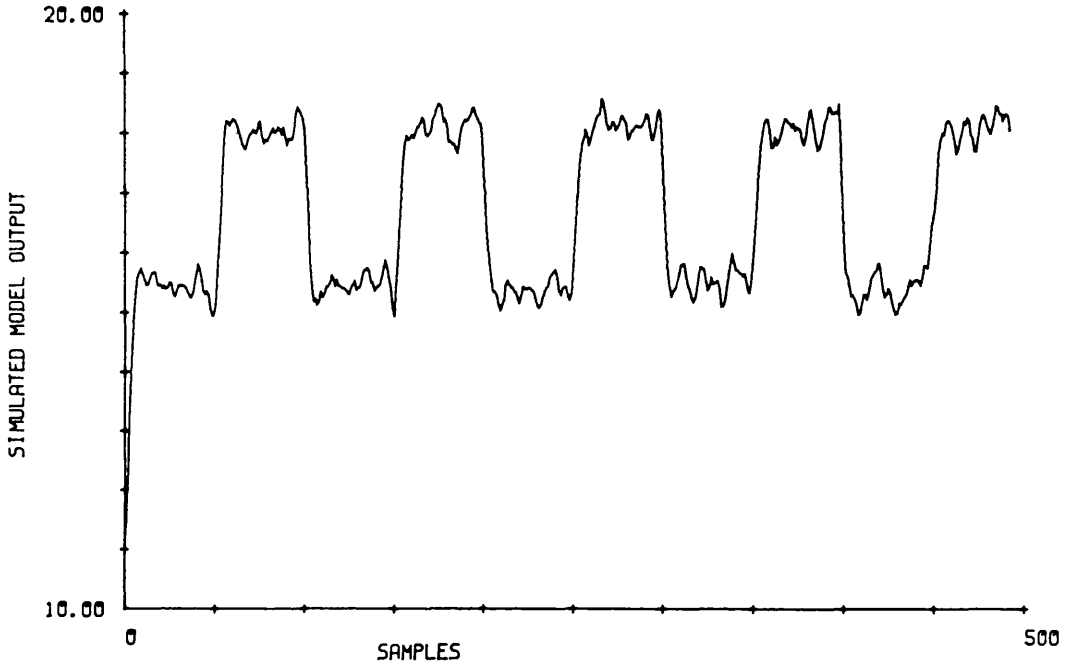
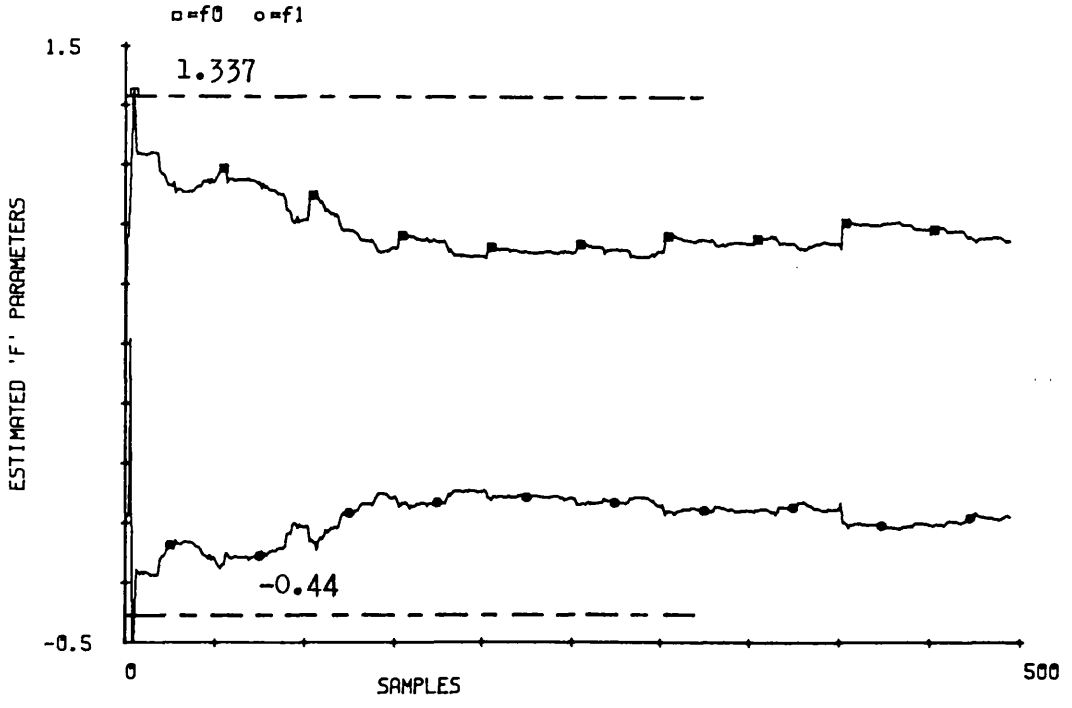


Fig. 4.5.1 Model output and 'F' parameter response for output given by Case 1



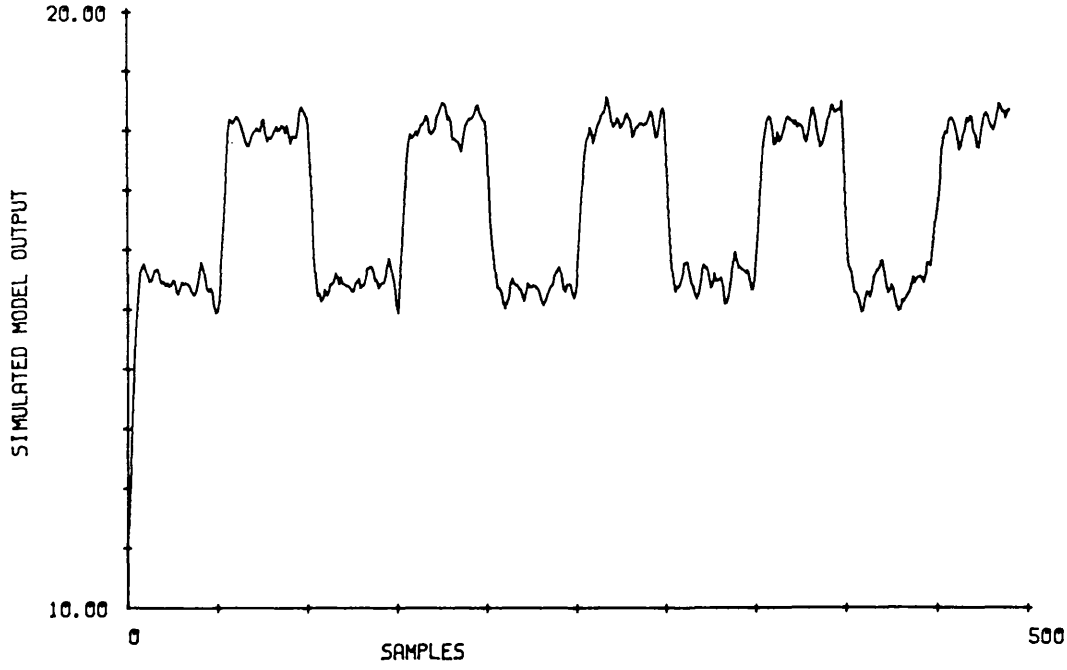
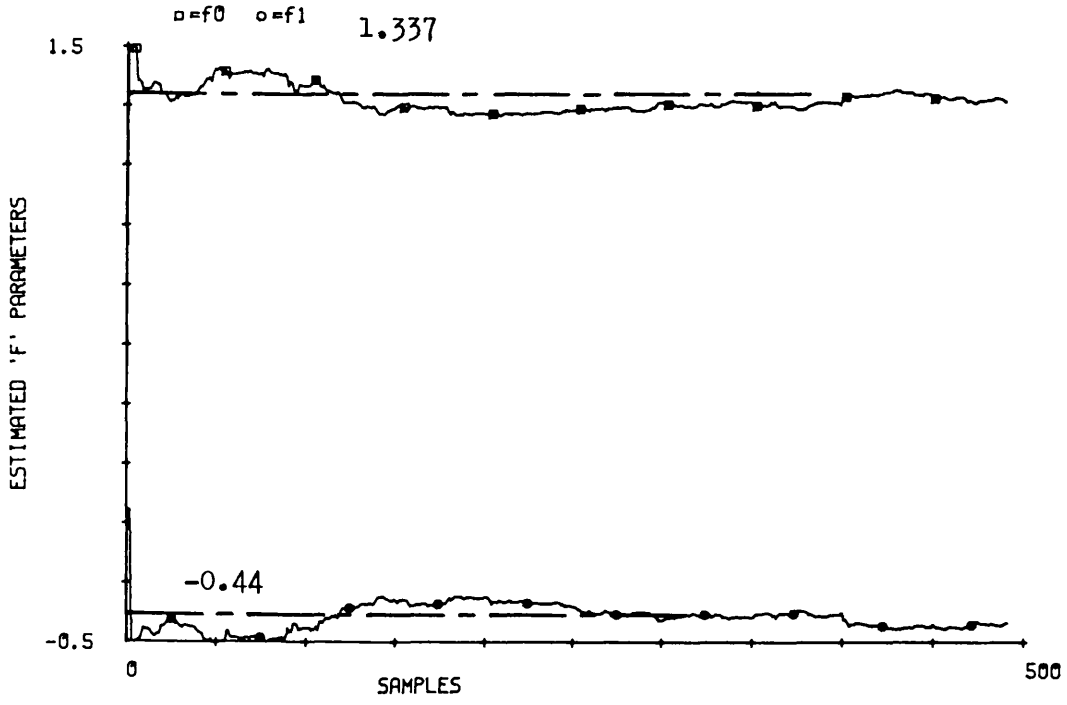


Fig. 4.5.2 Model output and 'F' parameter response for output given by Case 2



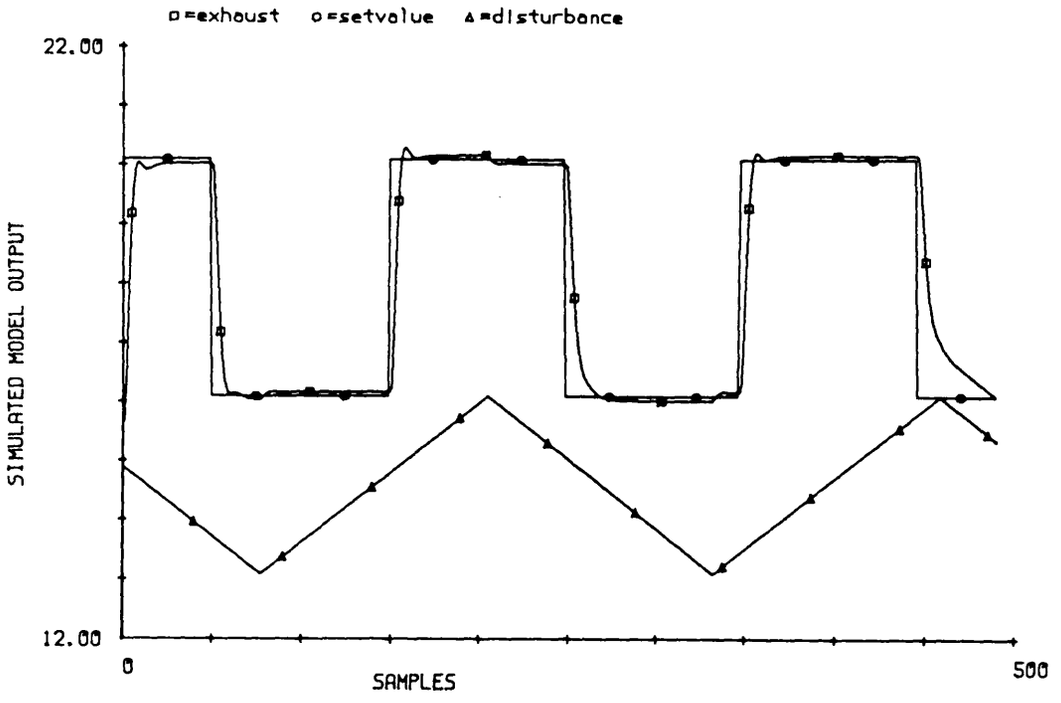


Fig. 4.5.3 Model output for ramped offset disturbance

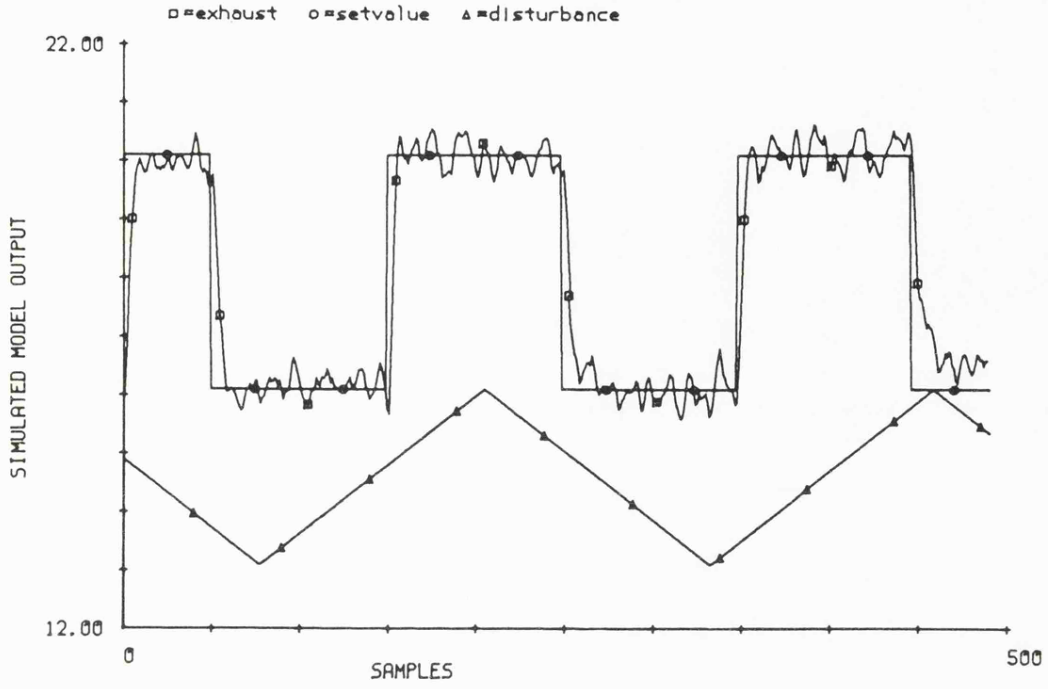
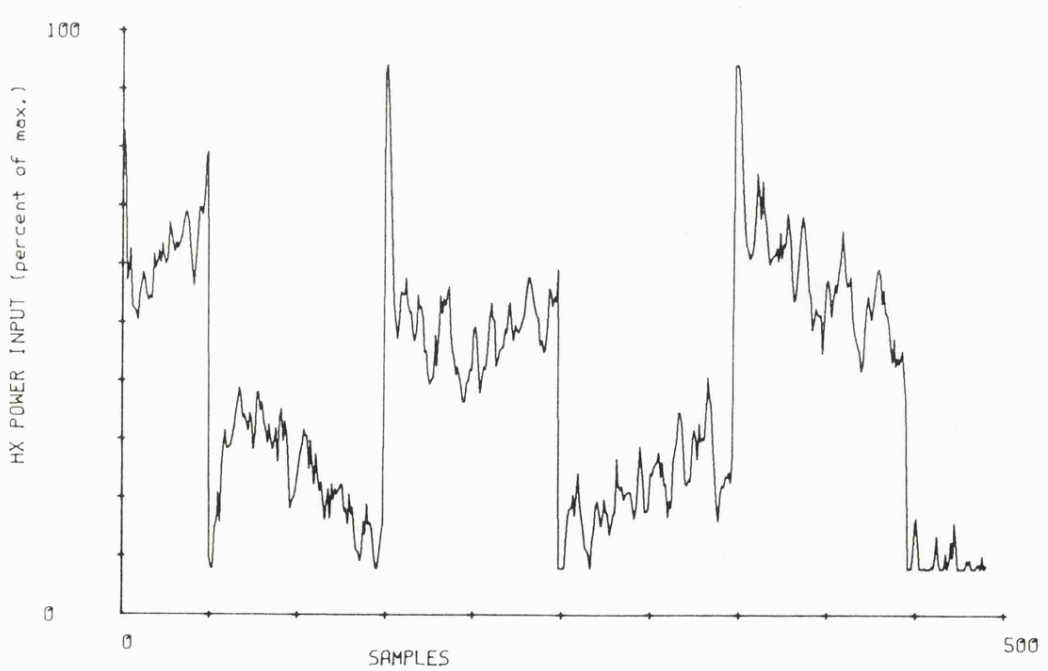


Fig. 4.5.4 Model output and heat exchanger input for ramped offset disturbance and nois y measurements



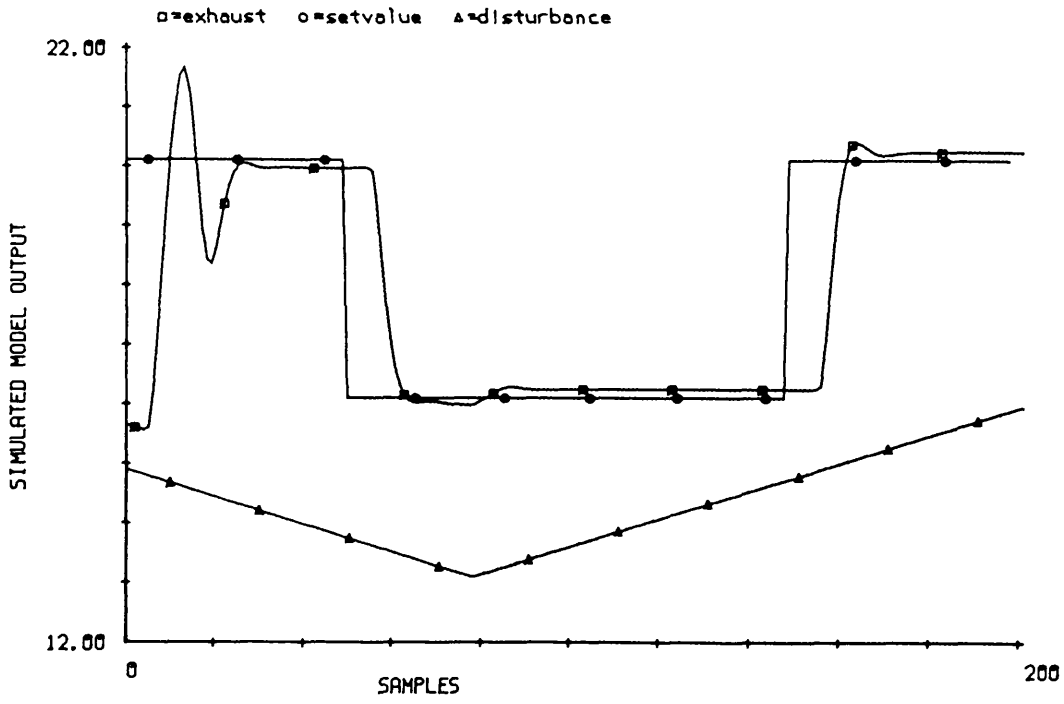
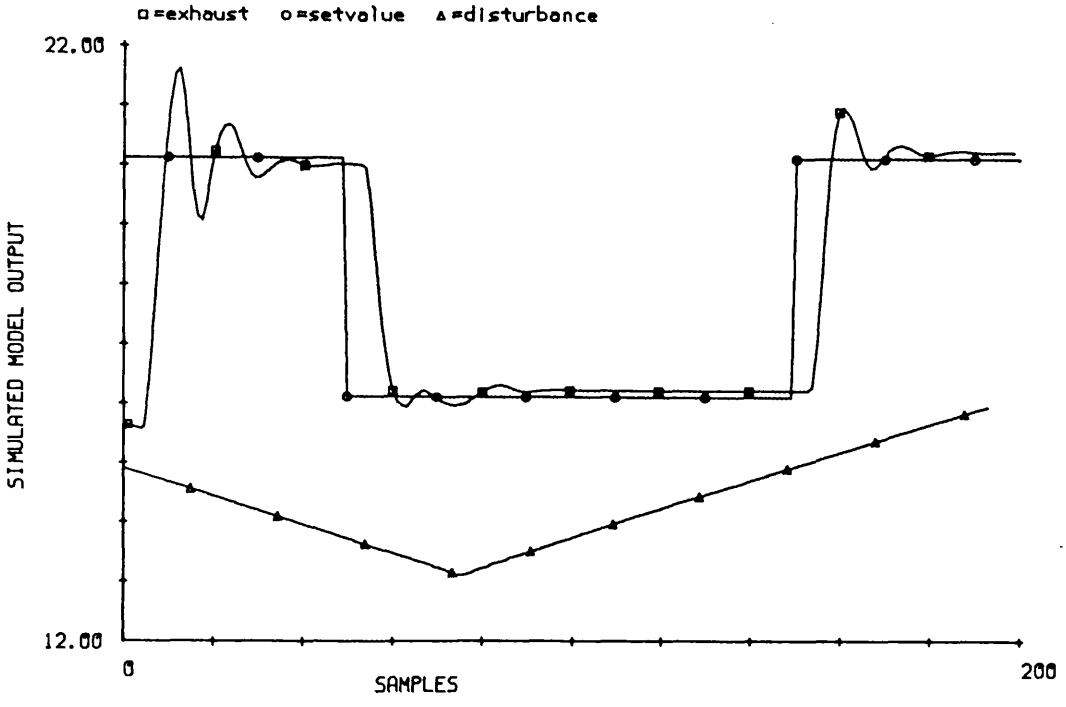


Fig. 4.5.5.a Model output for estimated delay 5 , actual delay 5

Fig. 4.5.5.b Model output for estimated delay 4 , actual delay 5



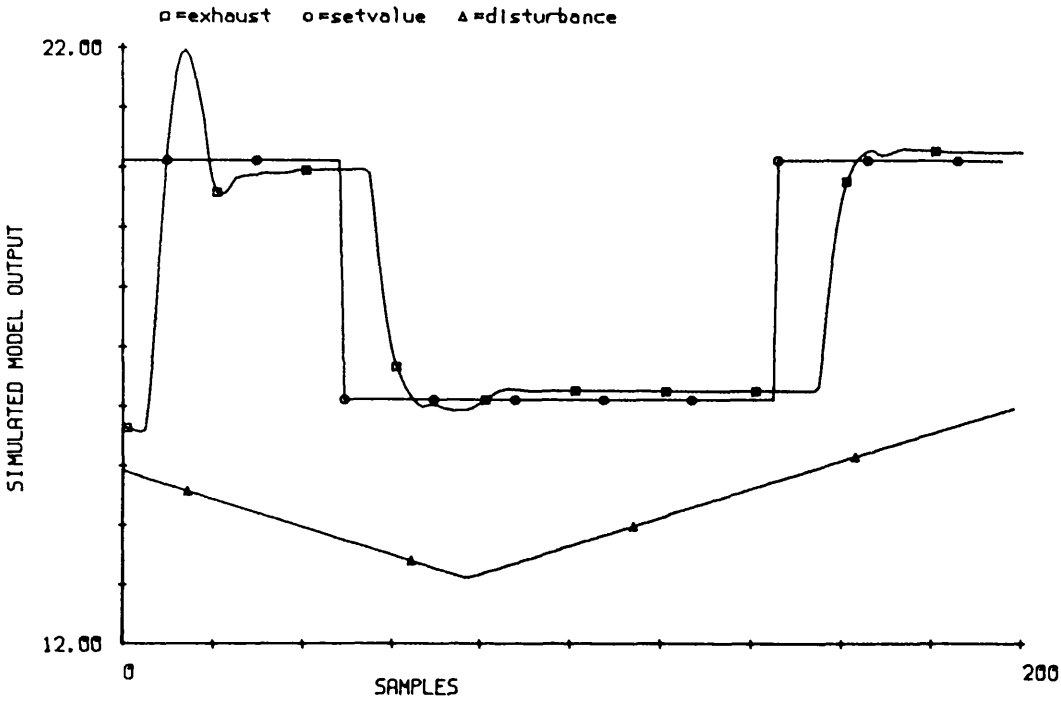


Fig. 4.5.5.c Model output for estimated delay 6 , actual delay 5

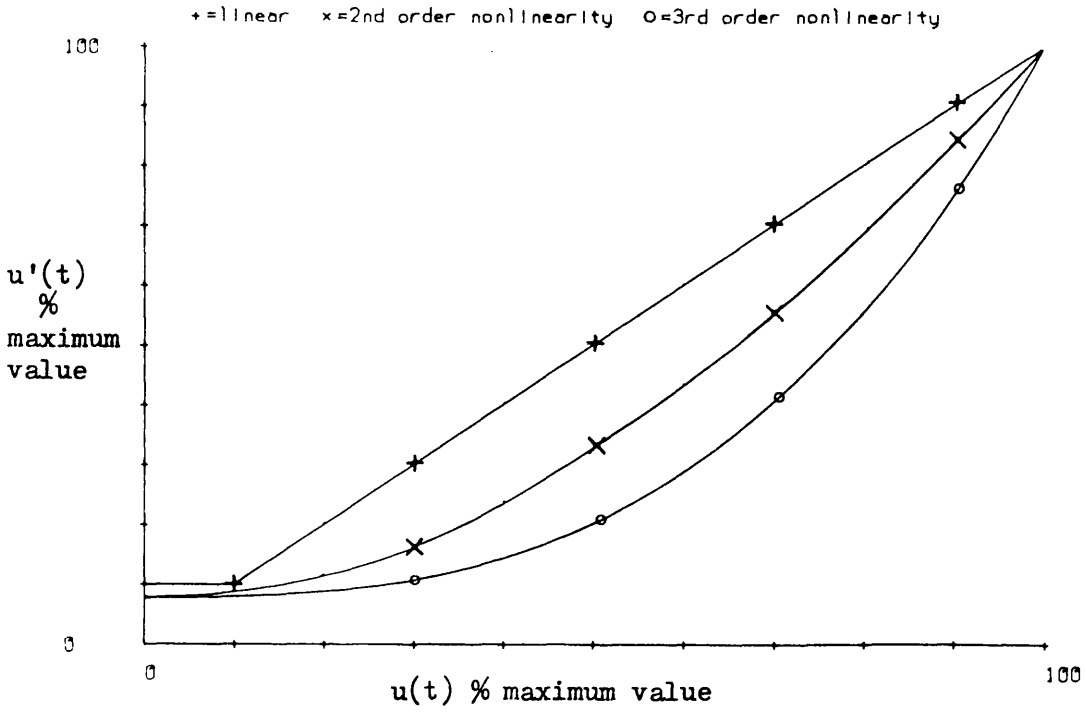


Fig. 4.5.6 Simulated plant forward gain characteristics

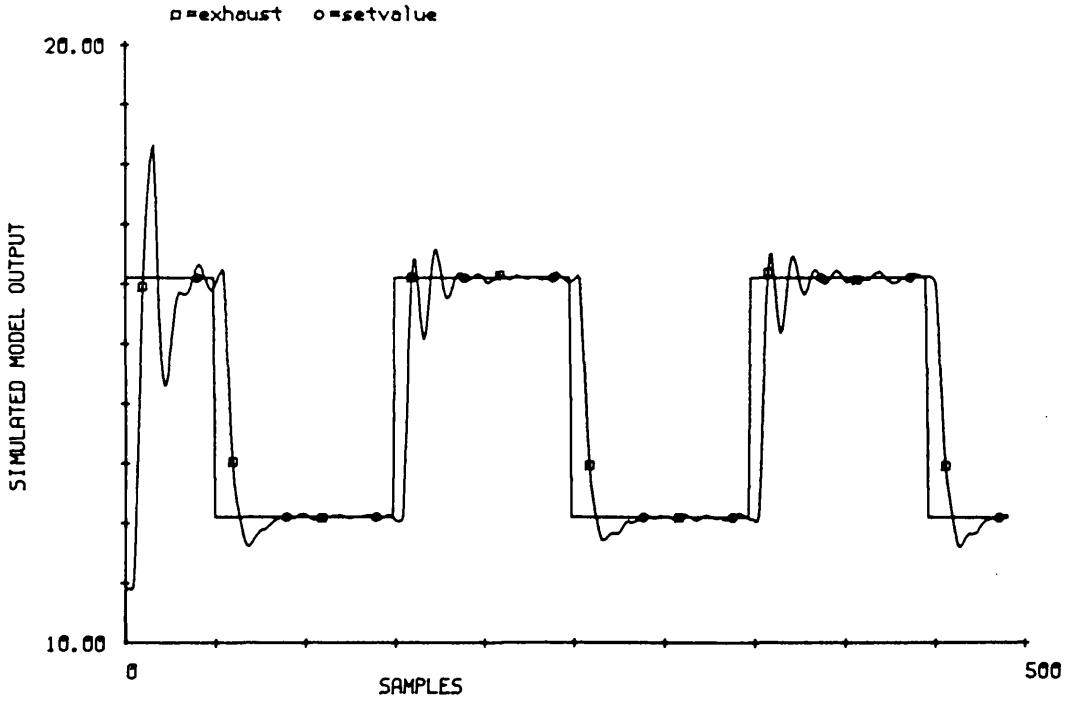
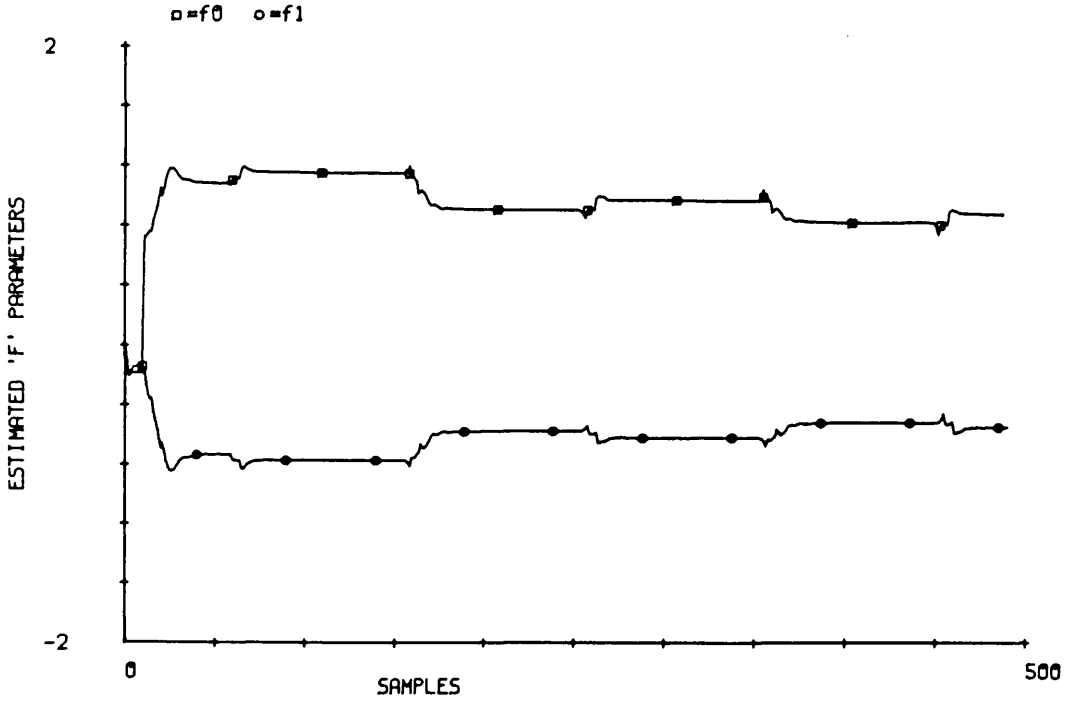


Fig. 4.5.7 Model output and 'F' parameter estimates for nonlinear plant and fixed forgetting factor



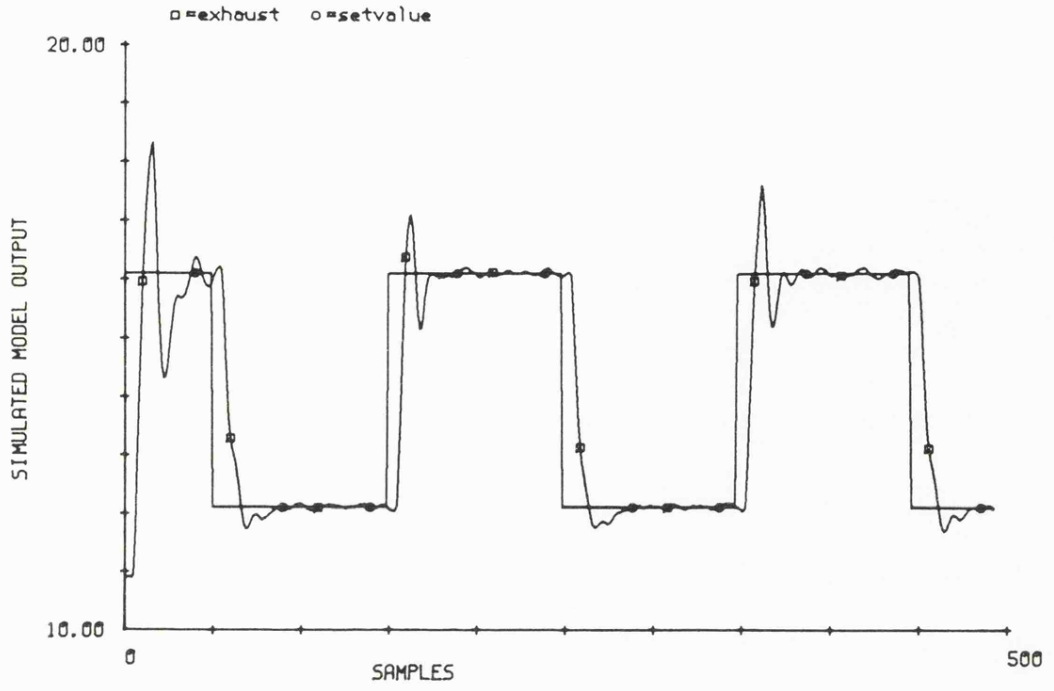
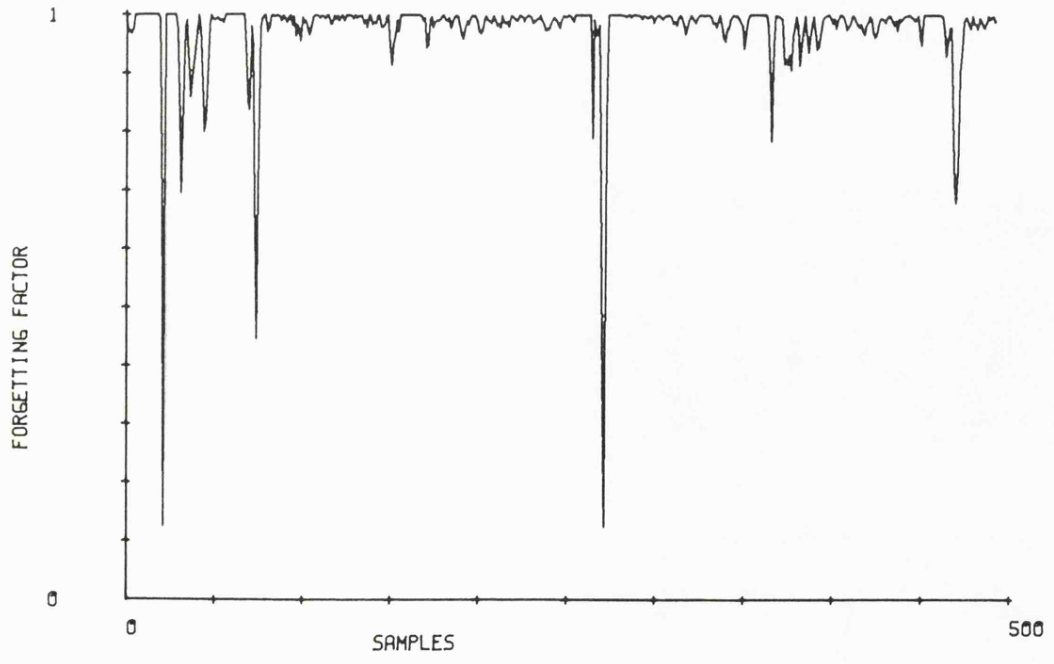


Fig. 4.5.8.a Model output and 'F' parameter estimates for nonlinear plant and variable forgetting factor



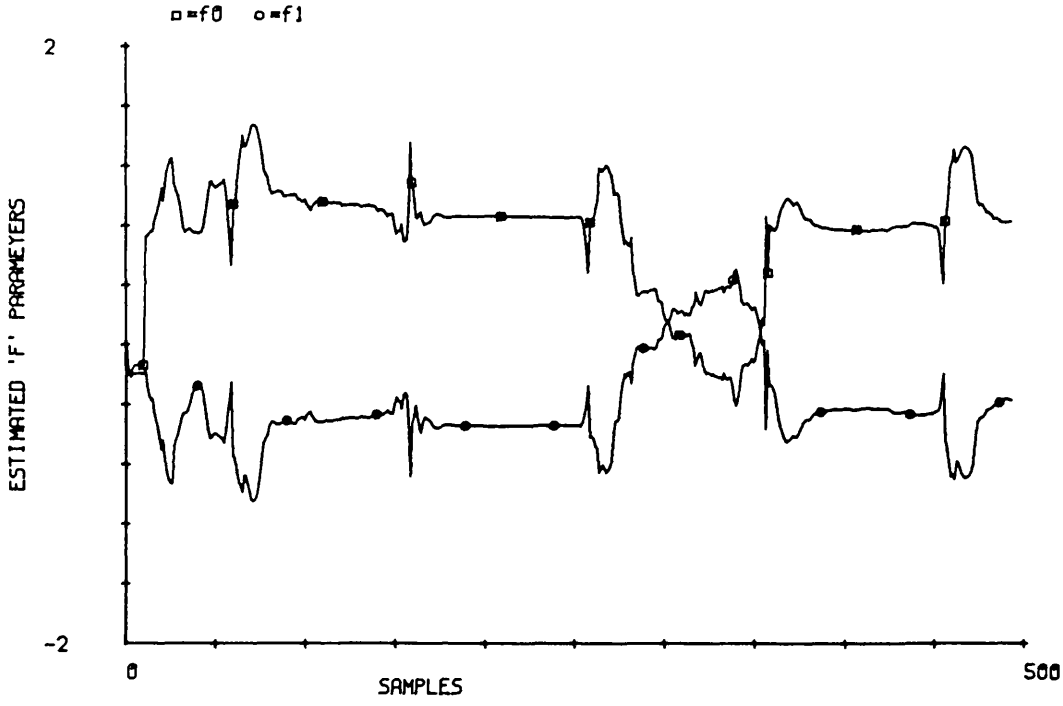


Fig. 4.5.8.b 'F' parameter estimates for variable forgetting factor

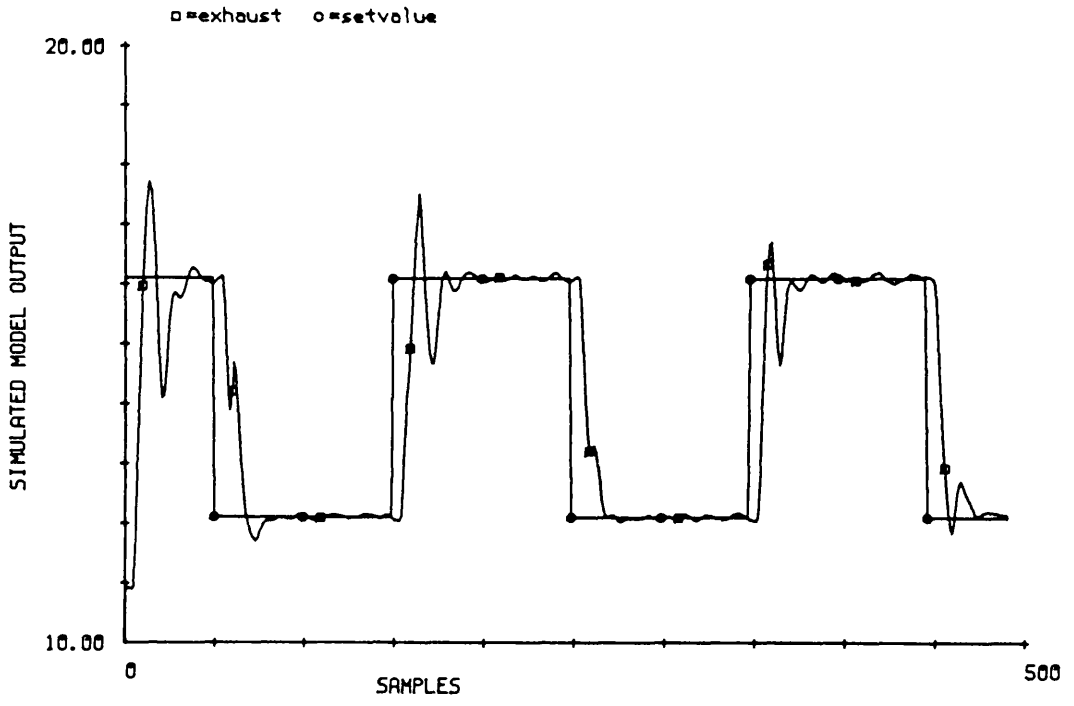
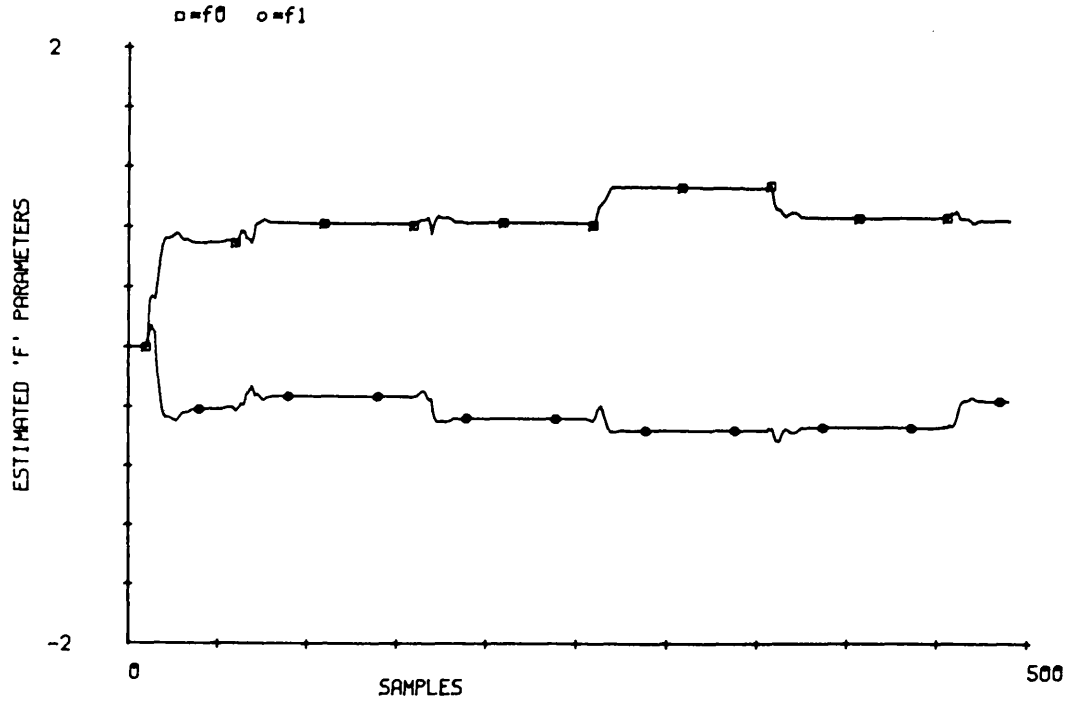


Fig. 4.5.9 Model output and 'F' parameter response for nonlinear plant and Covariance resetting



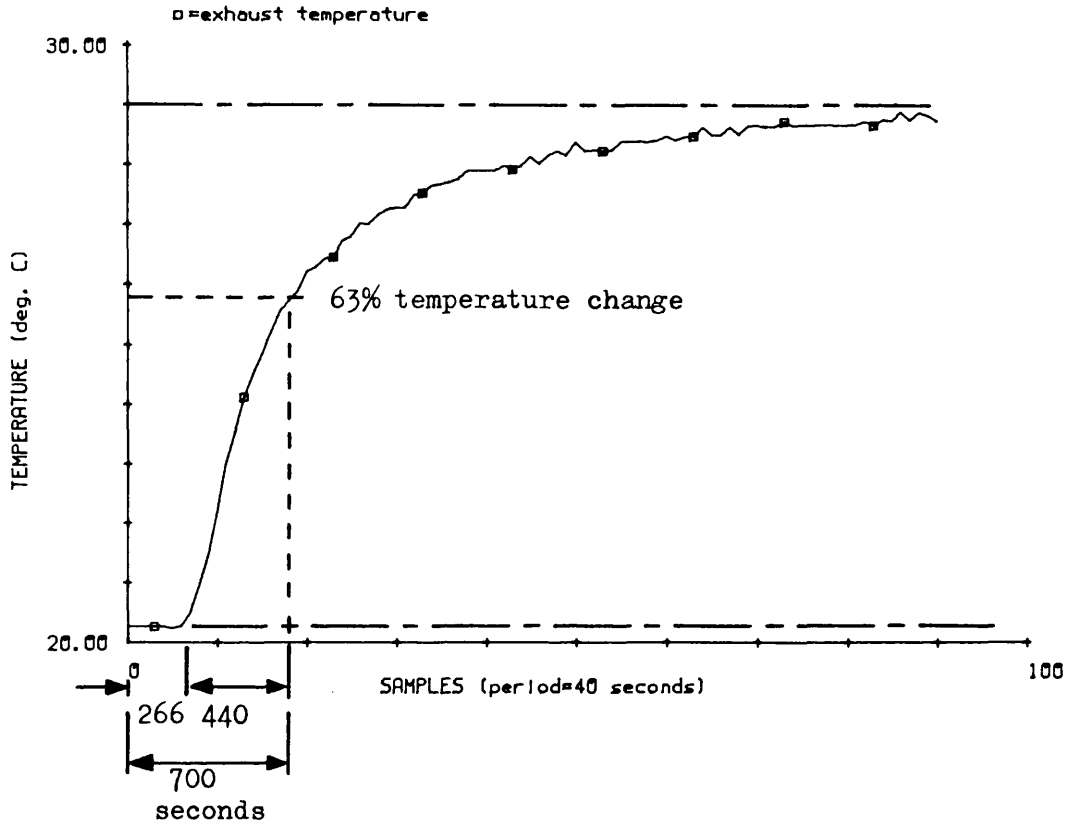


Fig. 5.1.1 Exhaust temperature response to 100% variation of heat exchanger input

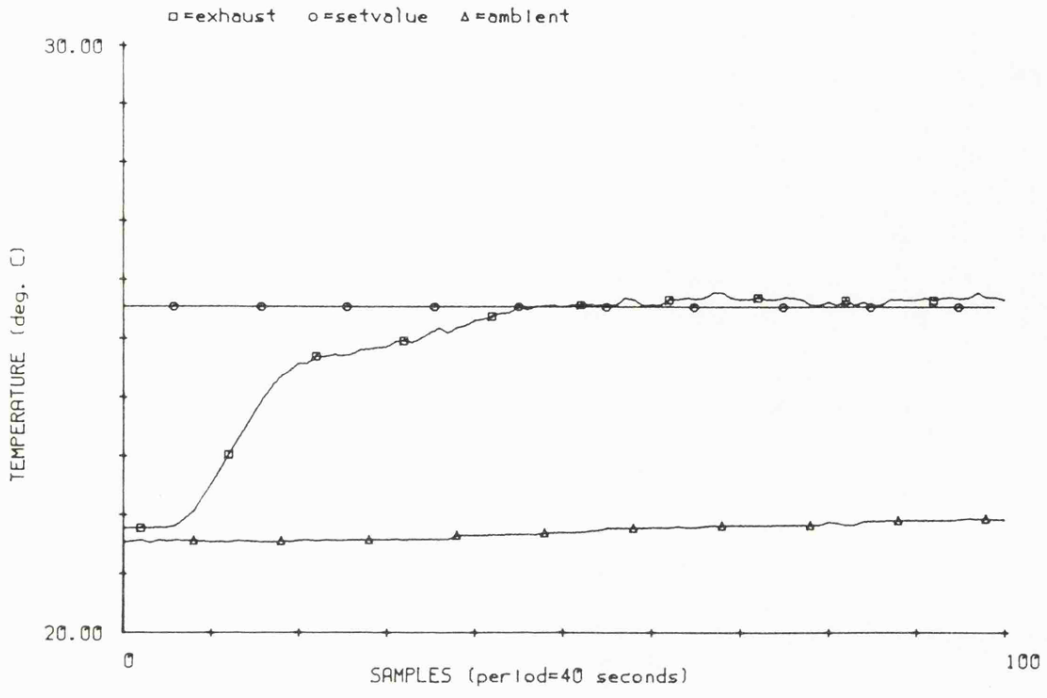
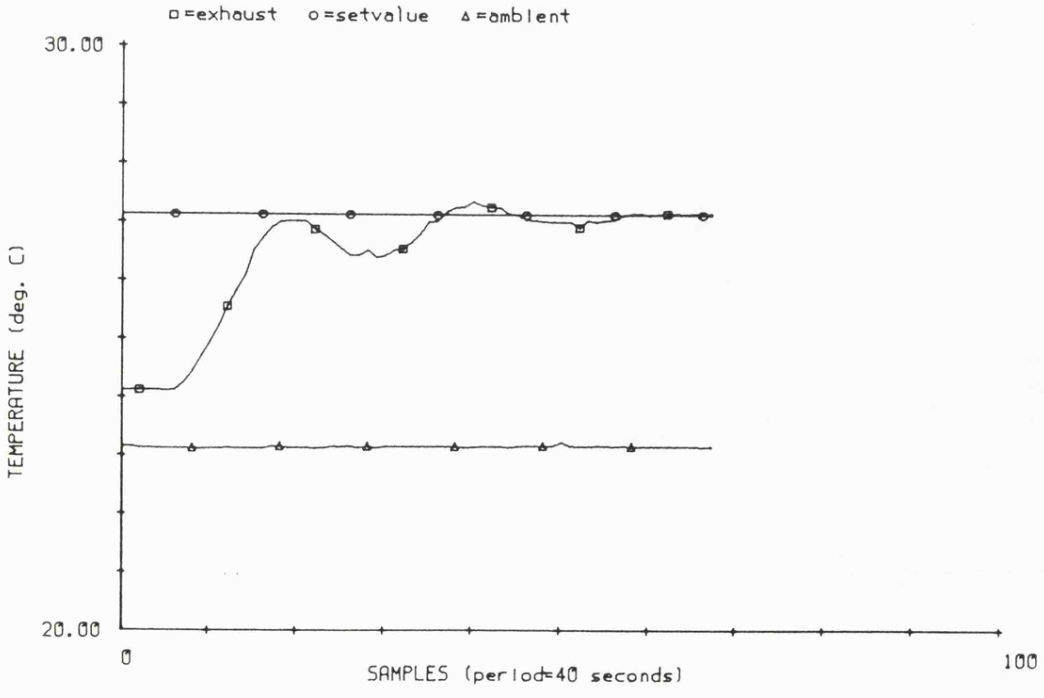


Fig. 5.1.2 Step response for control given by Eq'n 5.1.4

Fig. 5.1.3 Step response for control given by Eq'n 5.1.6



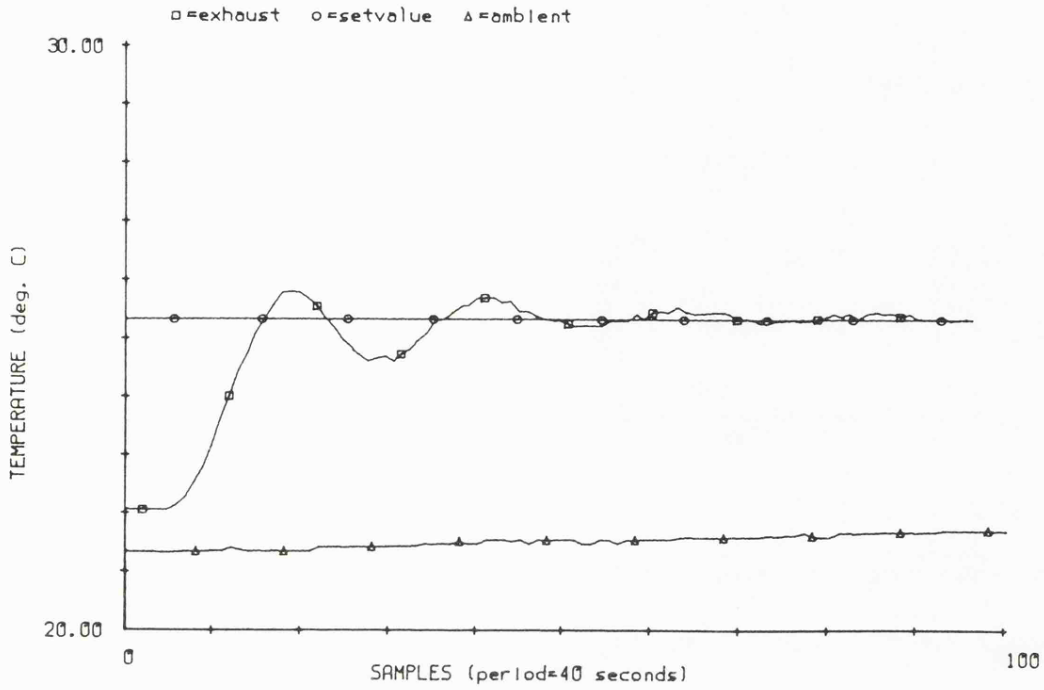
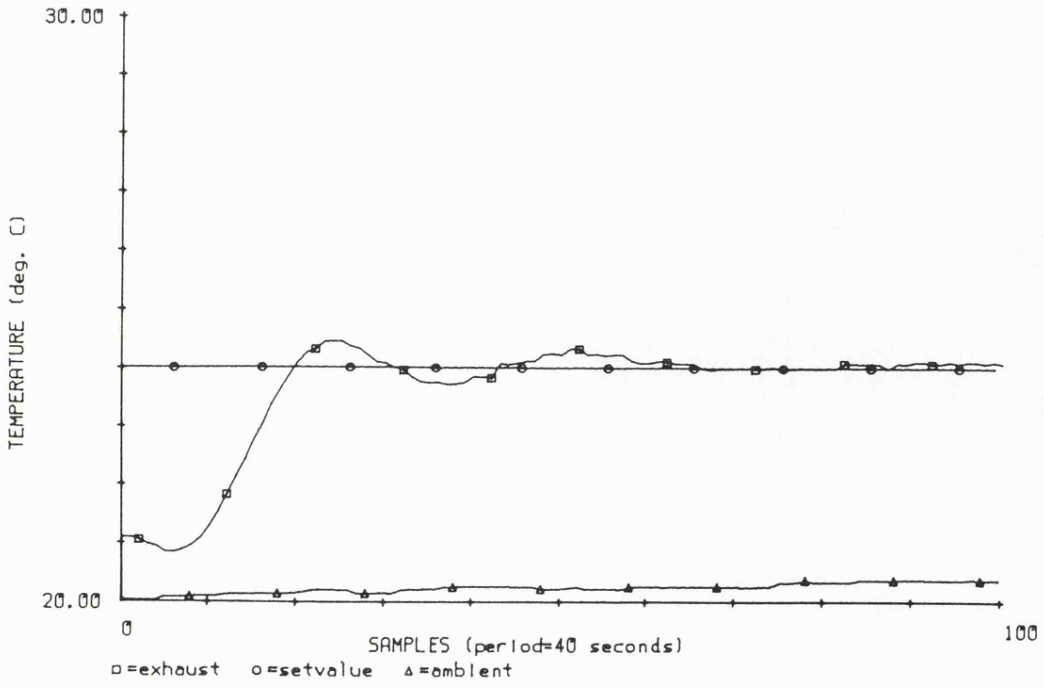


Fig. 5.1.4 Step response for control given by Eq'n 5.1.8

Fig. 5.1.5 Step response for control given by Eq'n 5.1.10



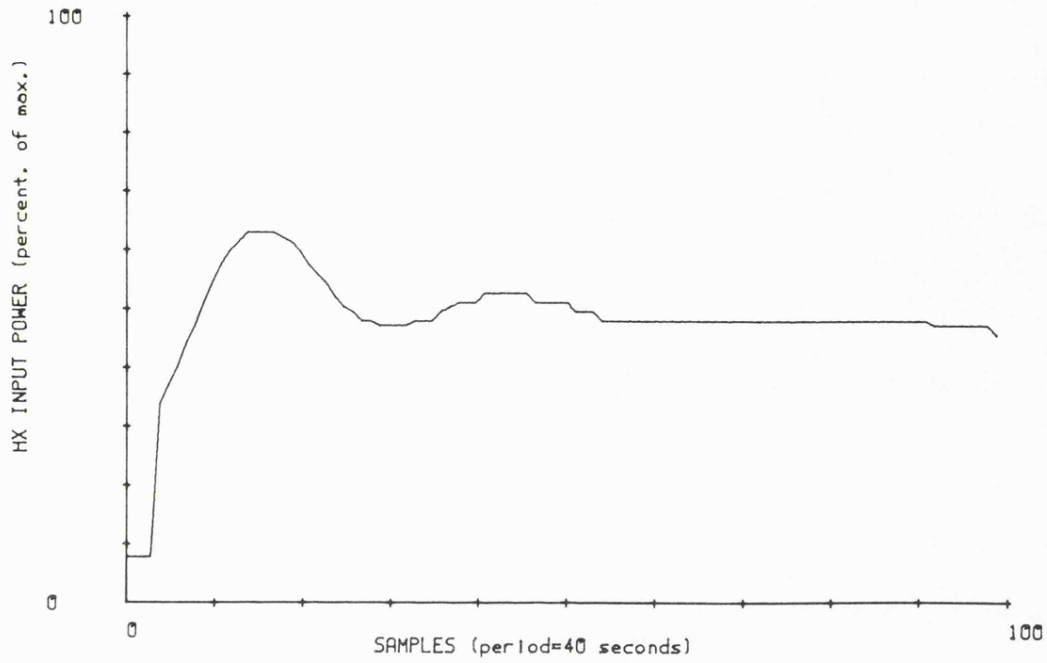


Fig. 5.1.6 Heat exchanger input for control given by Eq'n 5.1.10

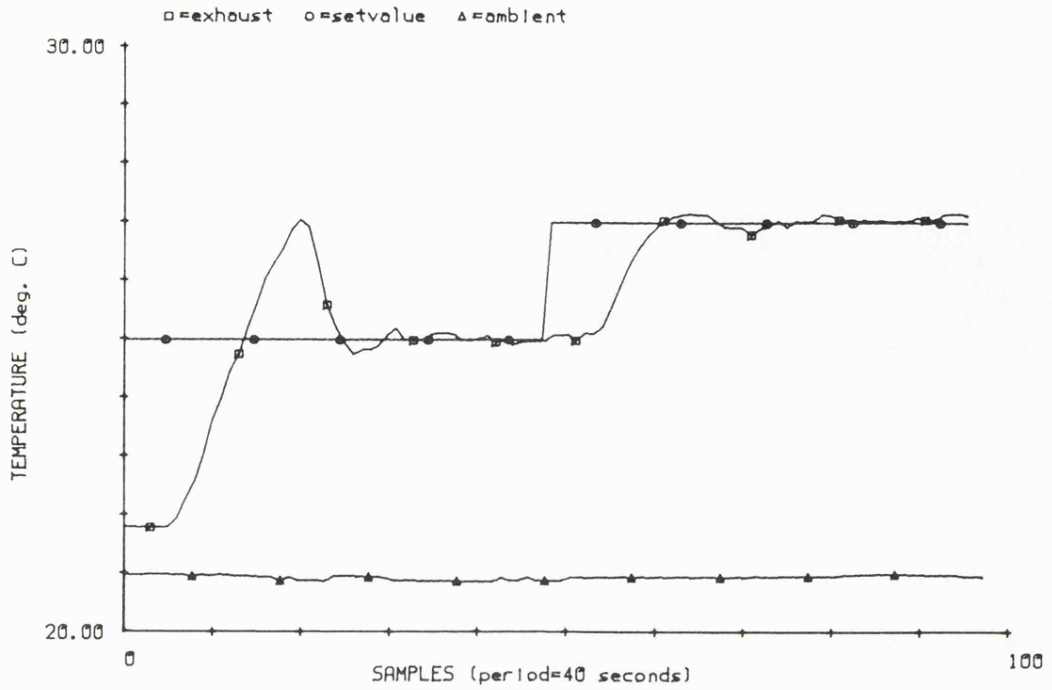
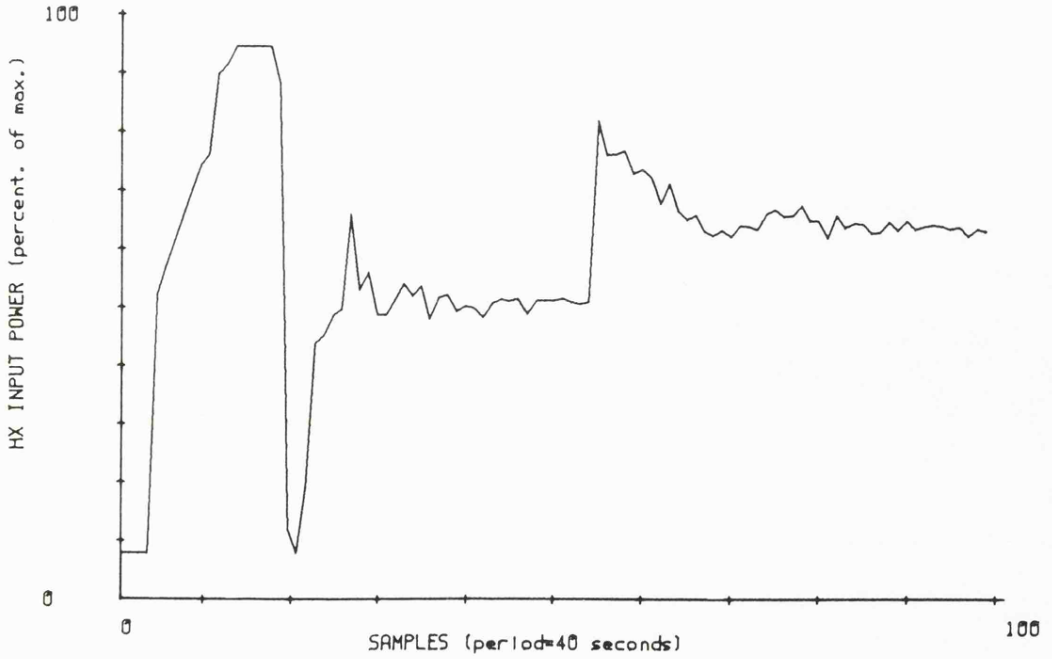


Fig. 5.1.7 Step response for self-tuning control

Fig. 5.1.8 Heat exchanger input for self-tuning control



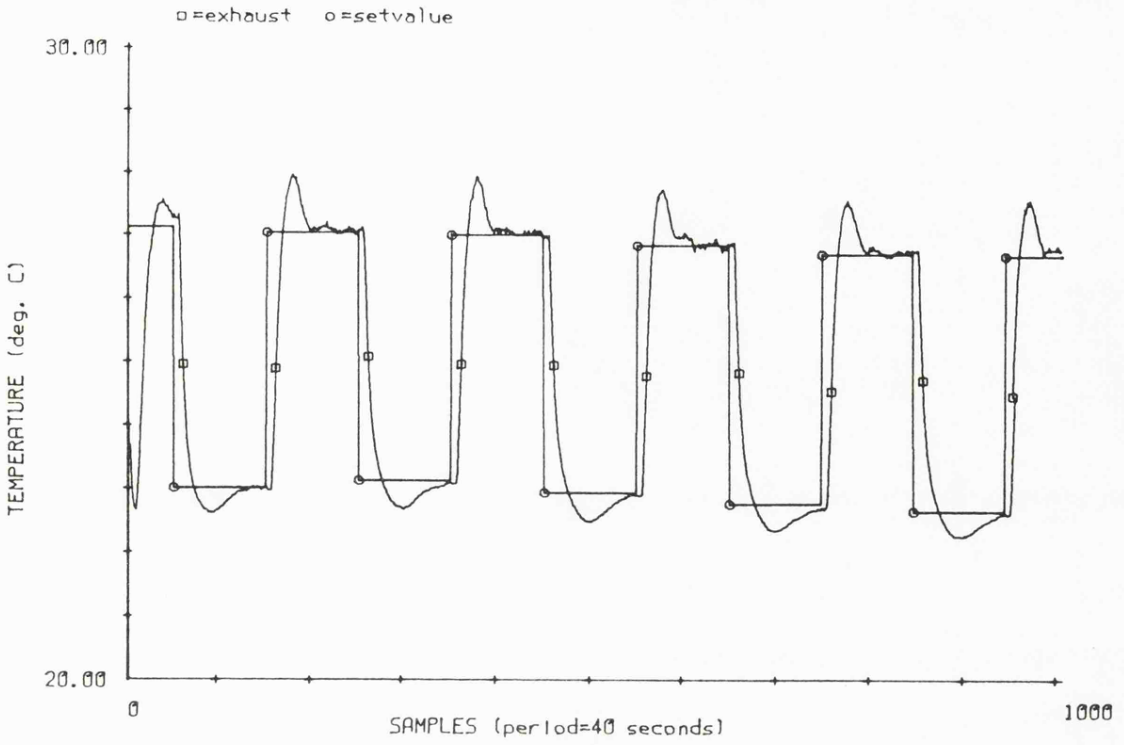
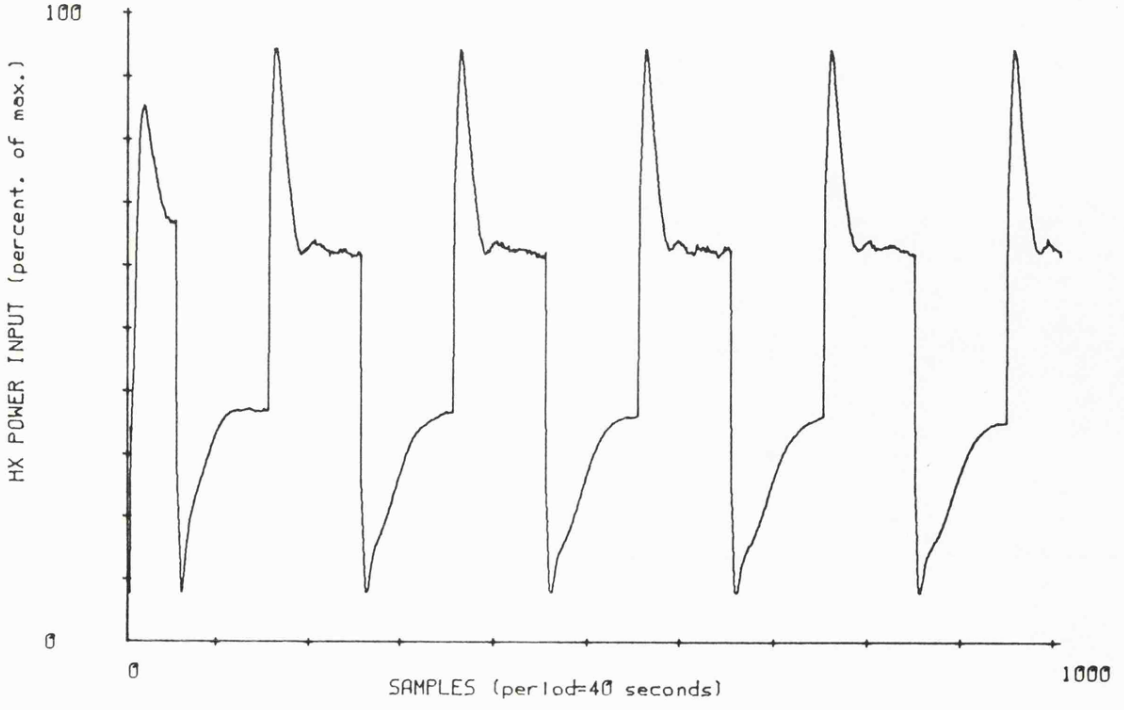


Fig. 5.2.1 Exhaust temperature response for conventional control

Fig. 5.2.2 Heat exchanger input for conventional control



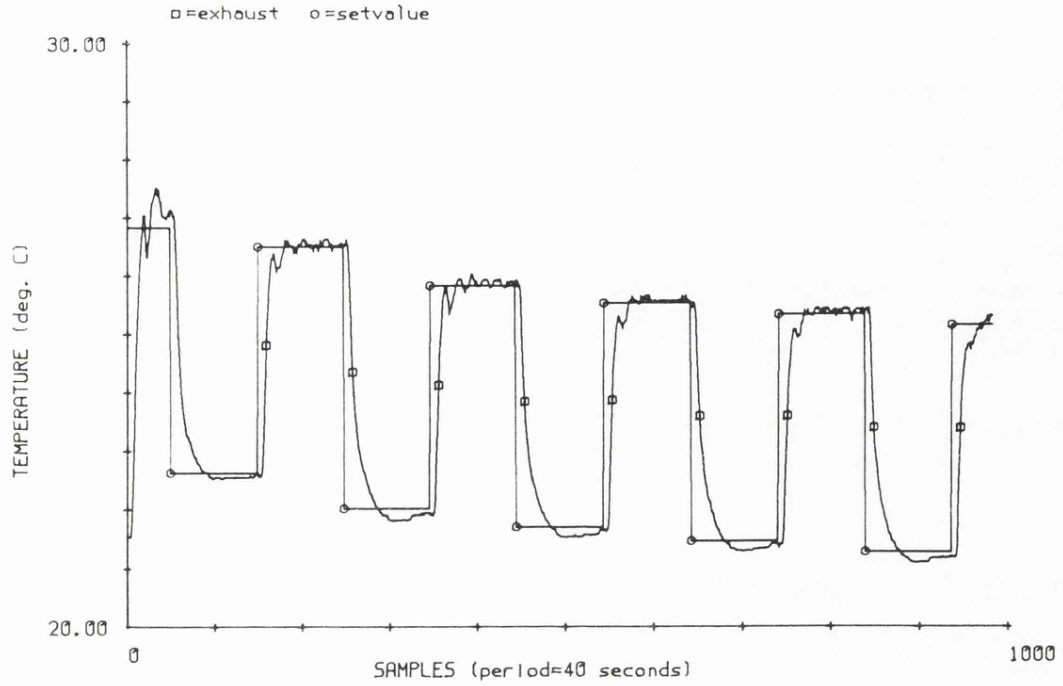
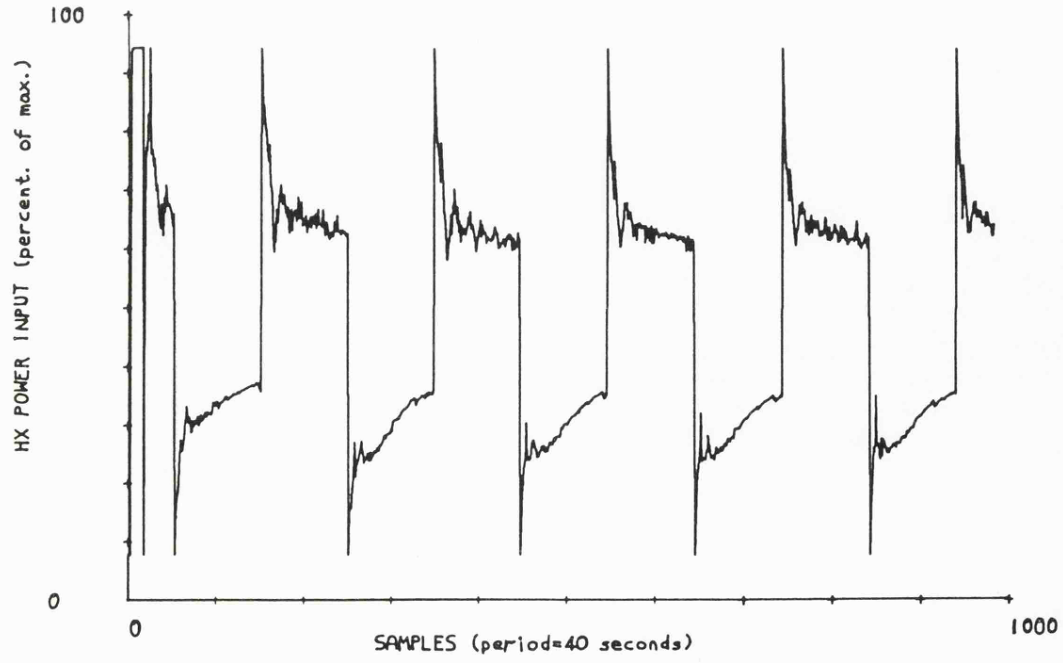


Fig. 5.2.3 Exhaust temperature response for self-tuning control Cost Function given by Eq'n 5.2.1

Fig. 5.2.4 Heat exchanger input for self-tuning control



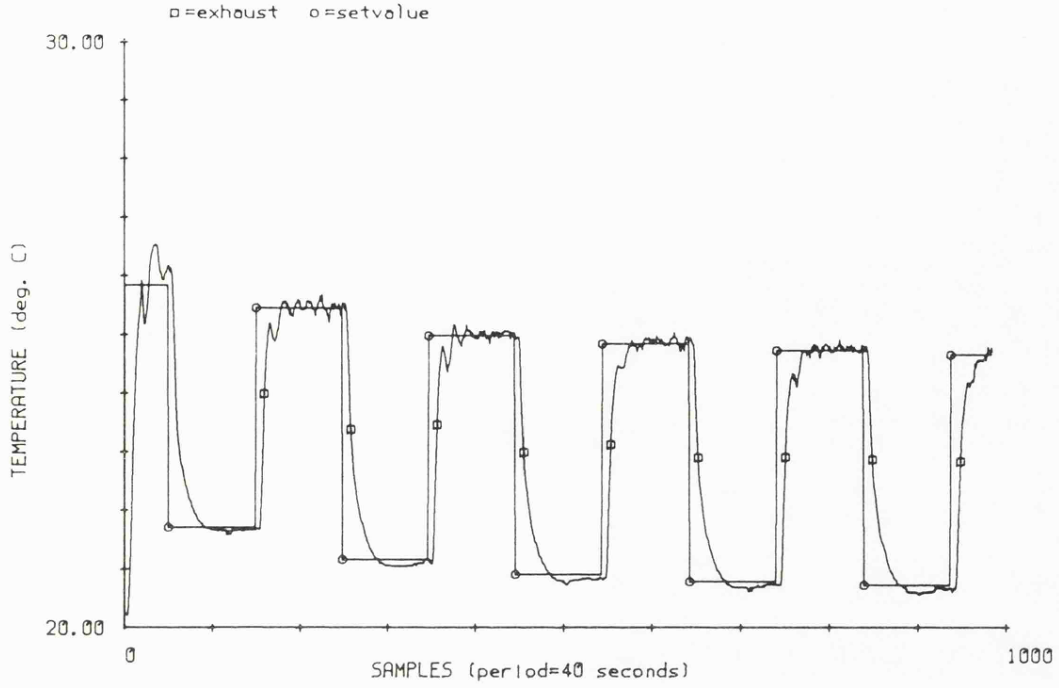
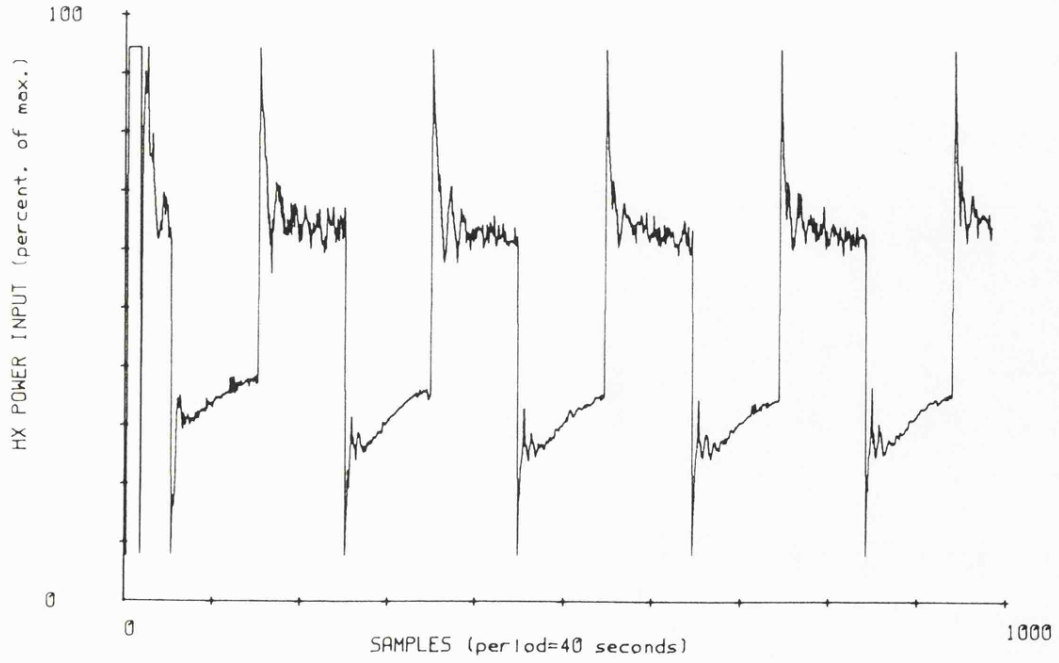


Fig. 5.2.5 Exhaust temperature response for self-tuning control Cost Function given by Eq'n 5.2.2

Fig. 5.2.6 Heat exchanger input response for self-tuning control



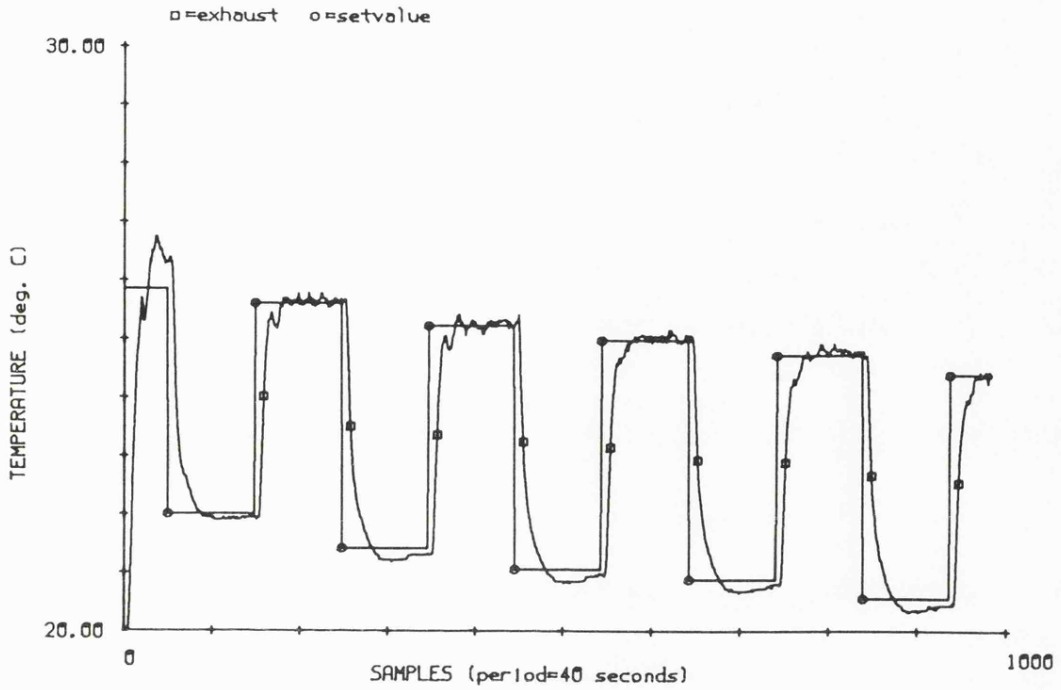
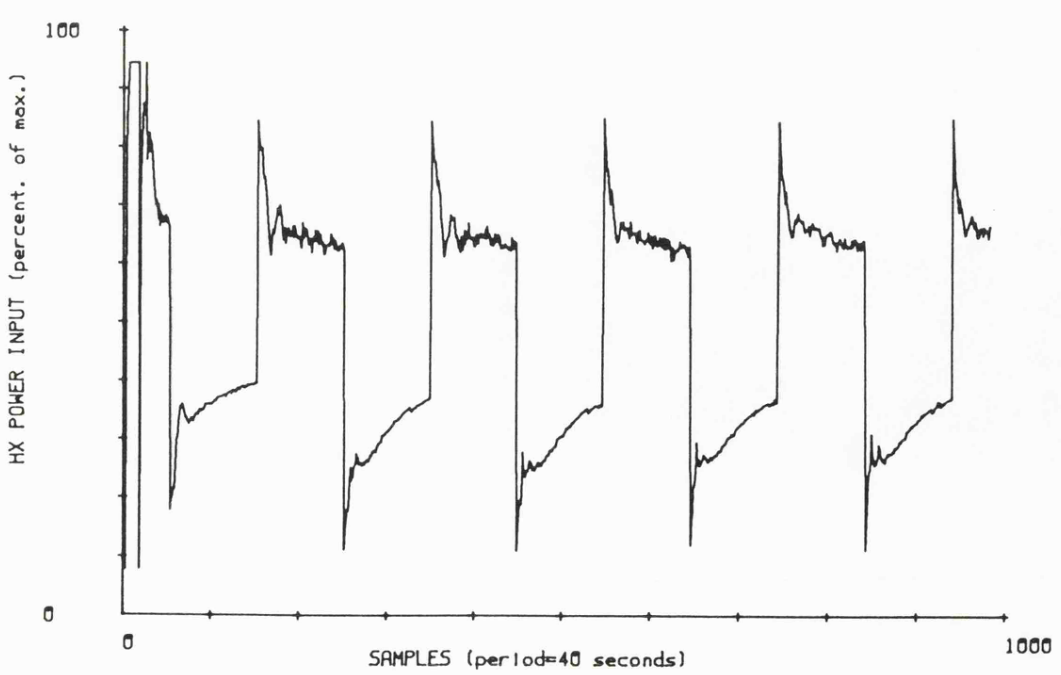


Fig. 5.2.7 Exhaust temperature response for self-tuning control Cost Function given by Eq'n 5.2.3

Fig. 5.2.8 Heat exchanger input for self-tuning control



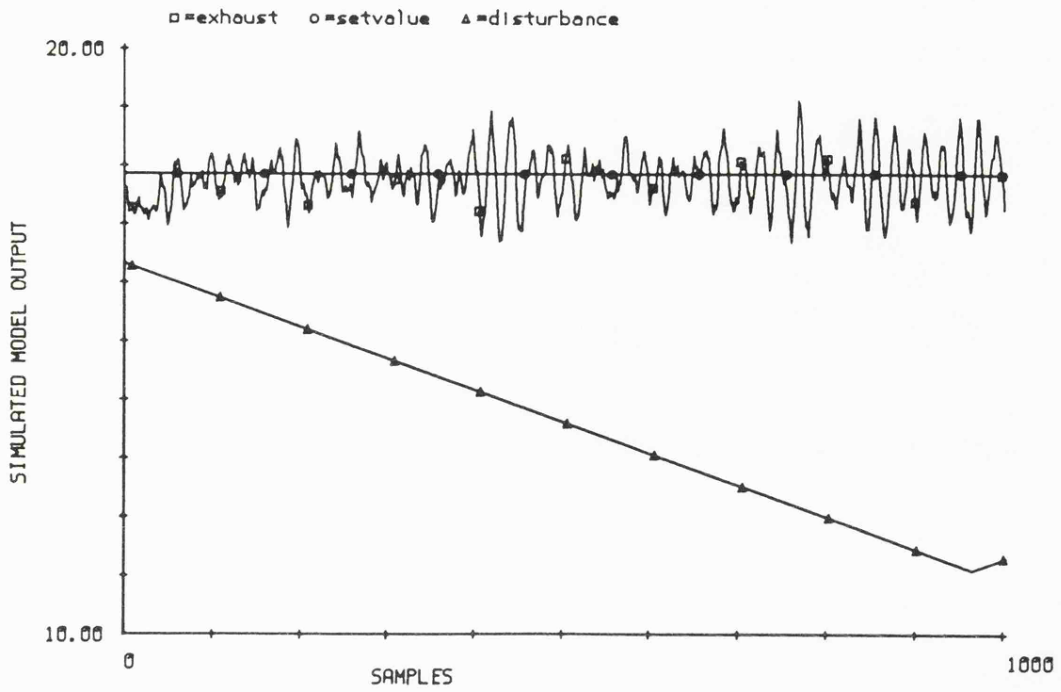
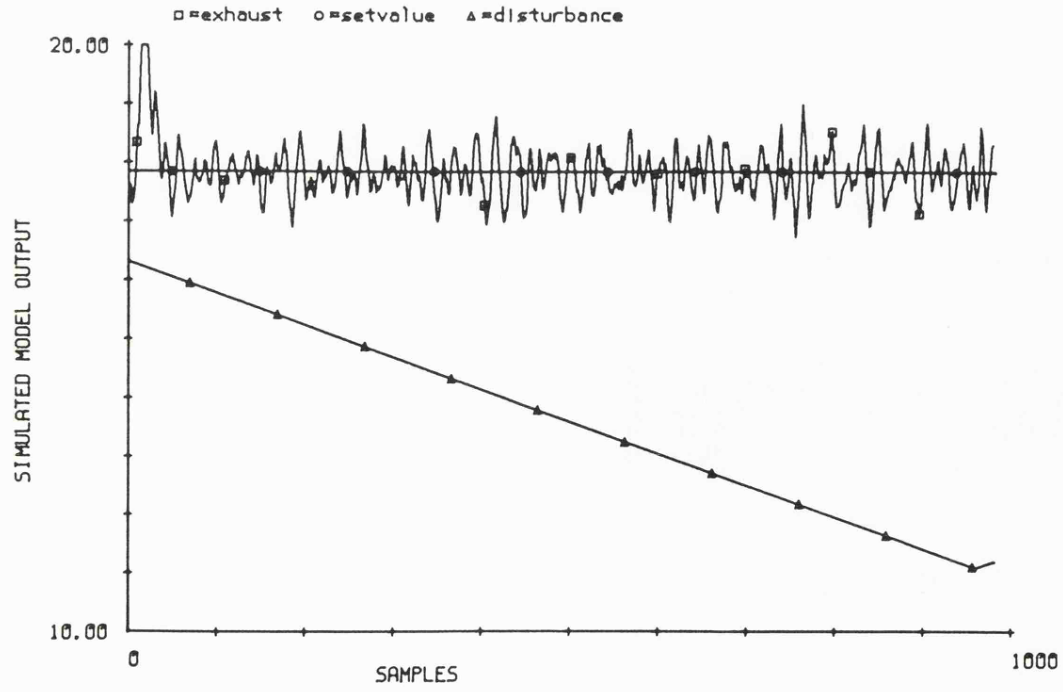


Fig. 5.2.9 Conventional control response to ramp plus noise disturbance

Fig. 5.2.10 Self-tuning control response to ramp plus noise disturbance



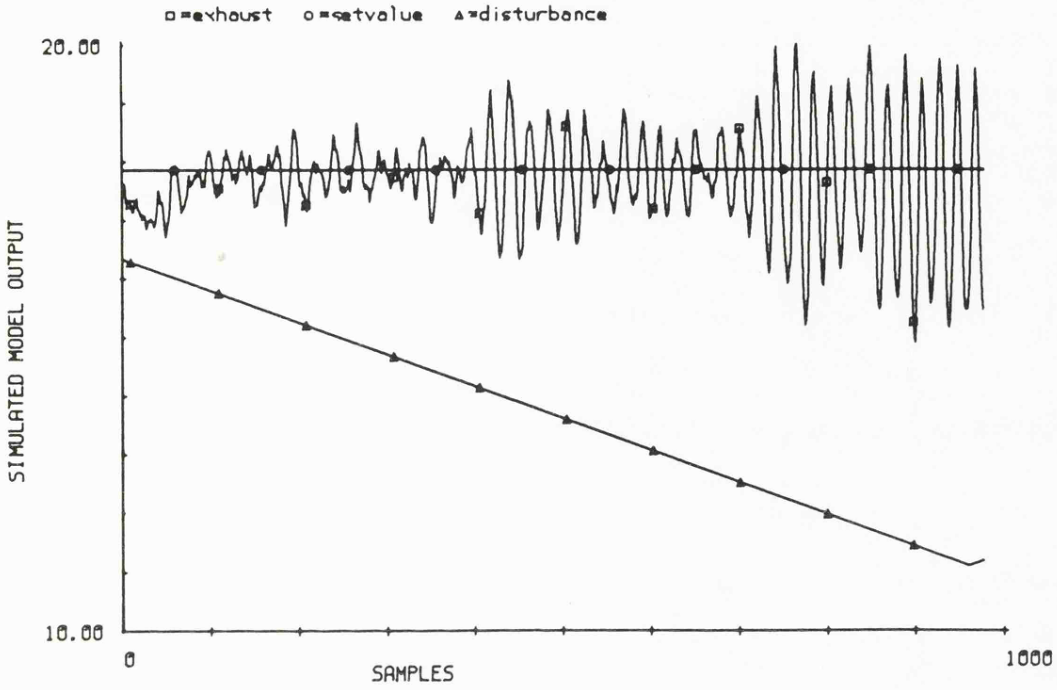
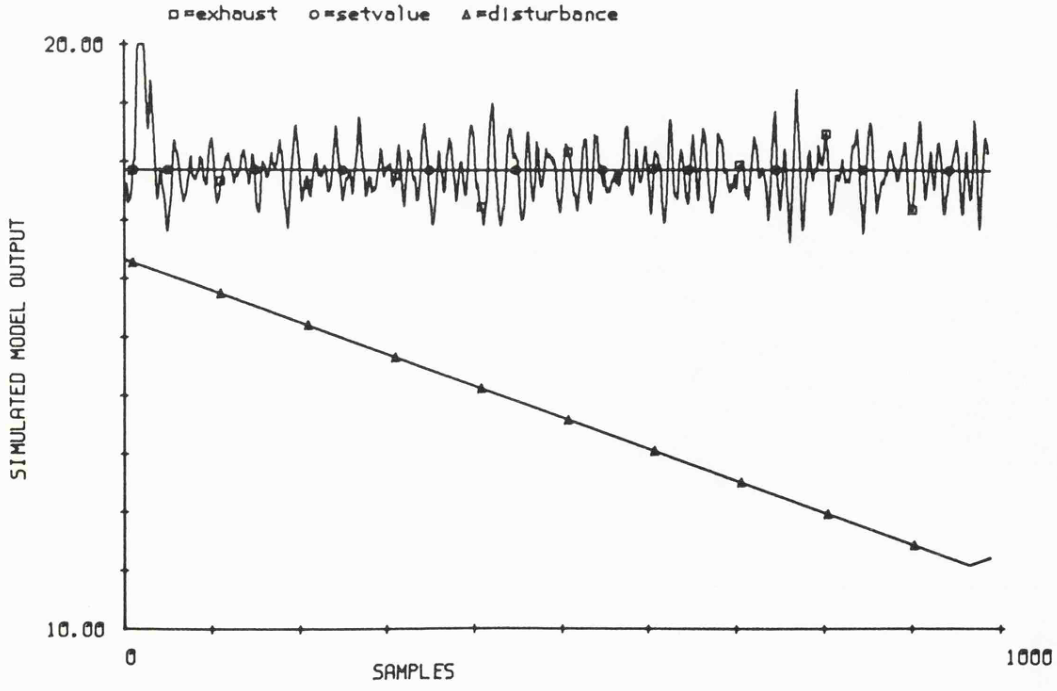


Fig. 5.2.11 Conventional control response to ramp plus noise disturbance

Fig. 5.2.12 Self-tuning control response to ramp plus noise disturbance



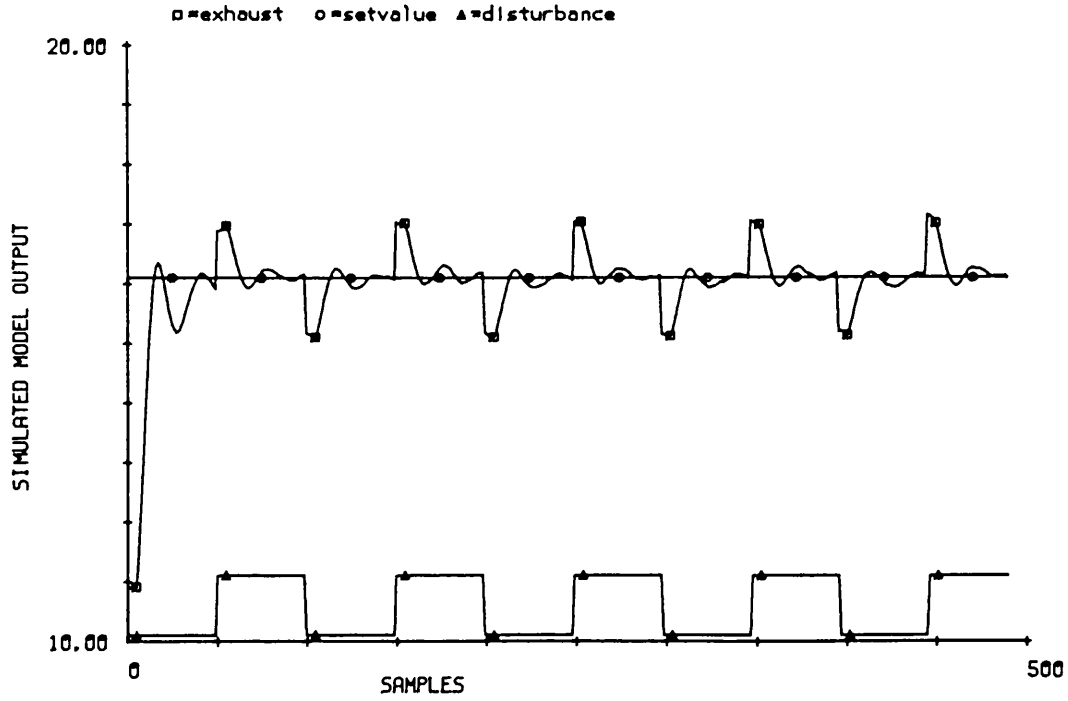
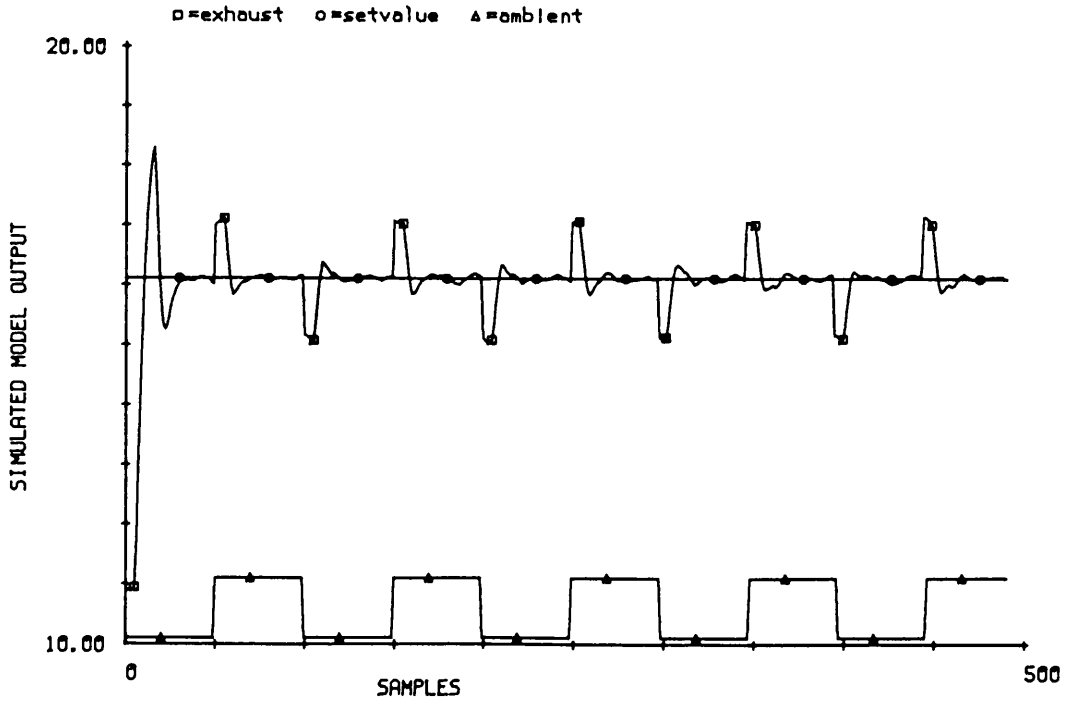


Fig. 5.2.13 Conventional control response to stepped disturbance

Fig. 5.2.14 Self-tuning control response to stepped disturbance



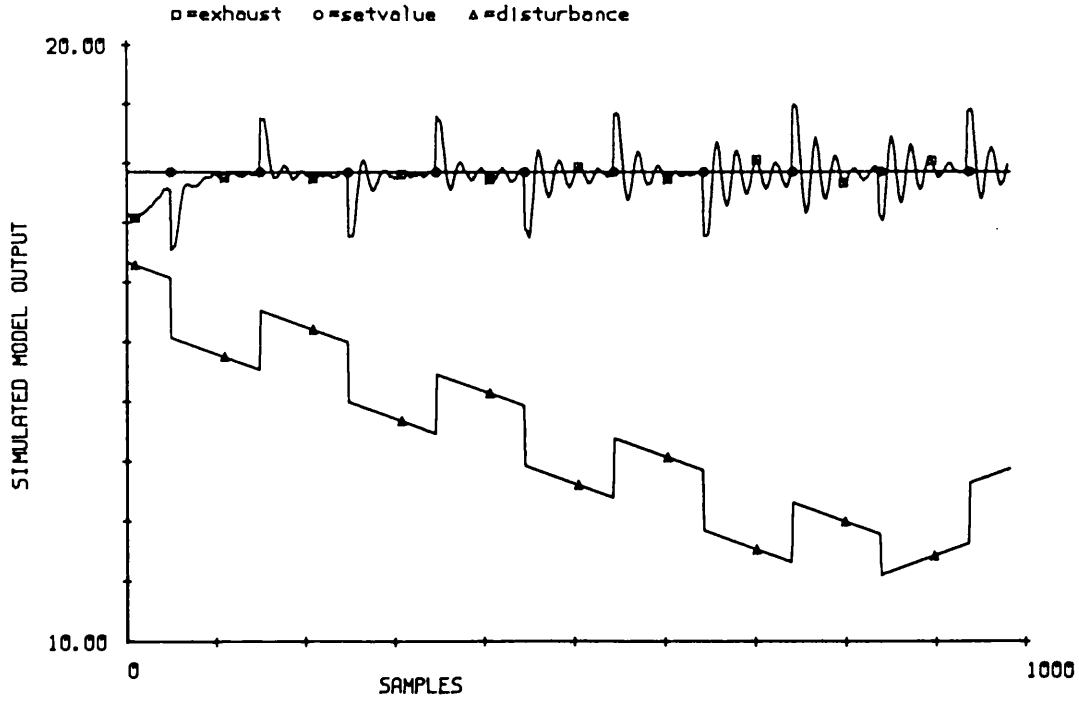
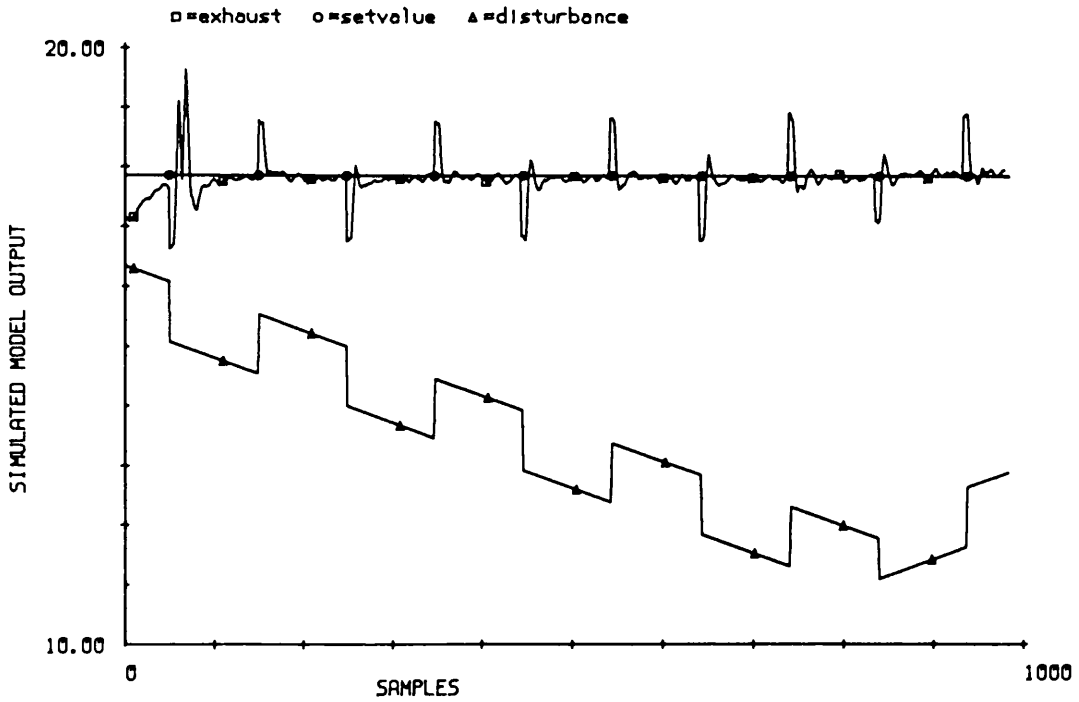


Fig. 5.2.15 Conventional control response to stepped plus ramp disturbance for nonlinear forward gain

Fig.5.2.16 Self-tuning control response to stepped plus ramp disturbance for nonlinear forward gain



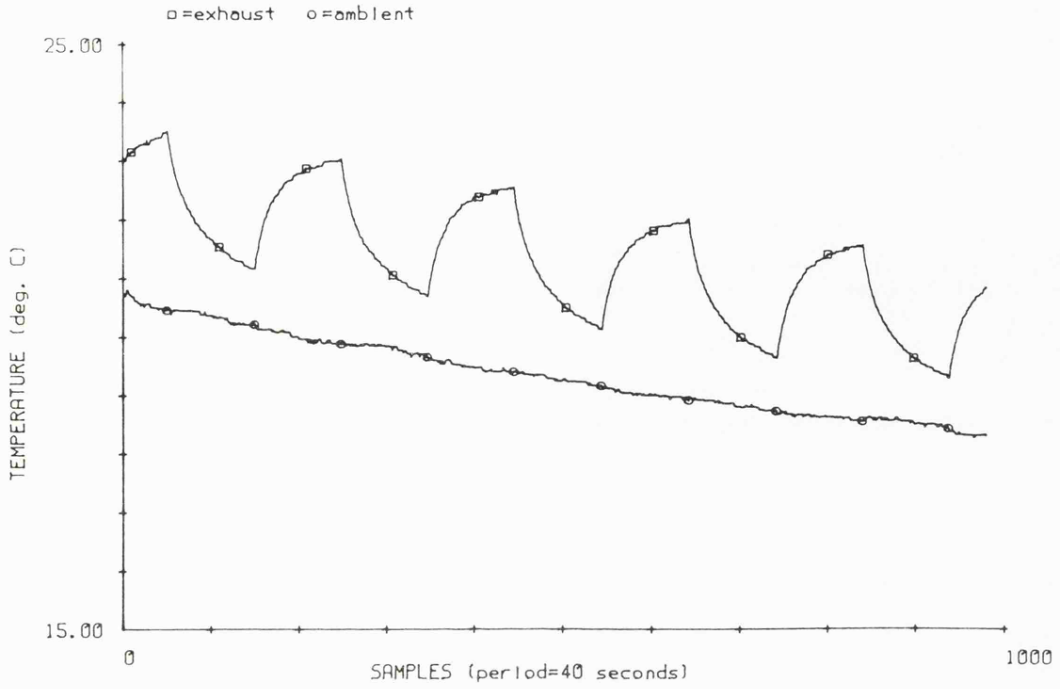


Fig. 5.2.17 Exhaust temperature response to 60 Watt radiant disturbance

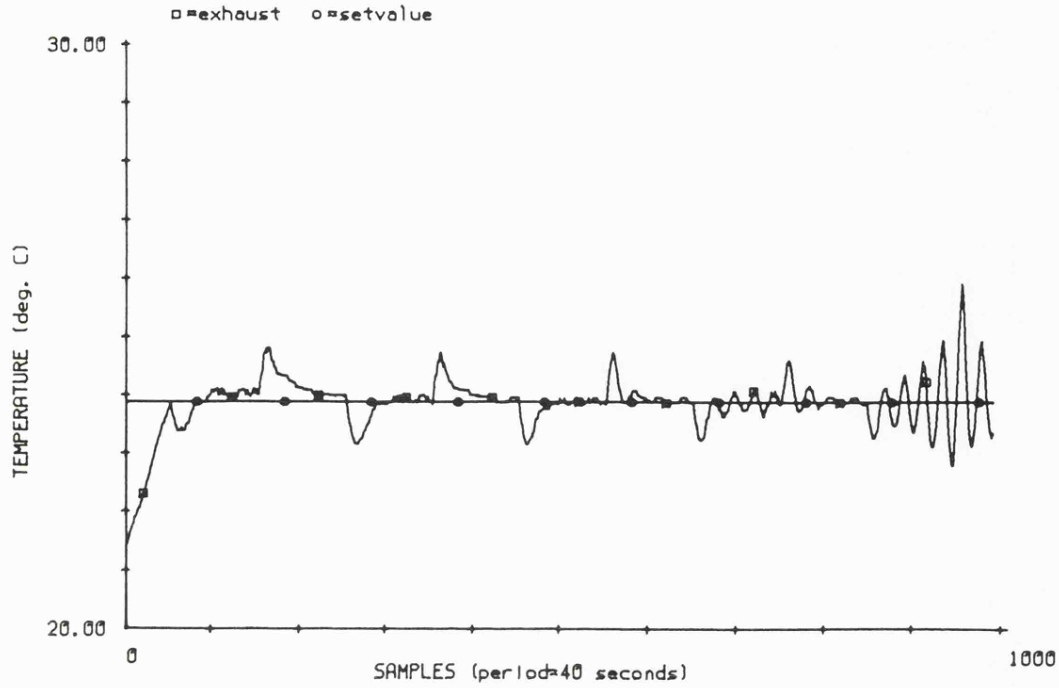
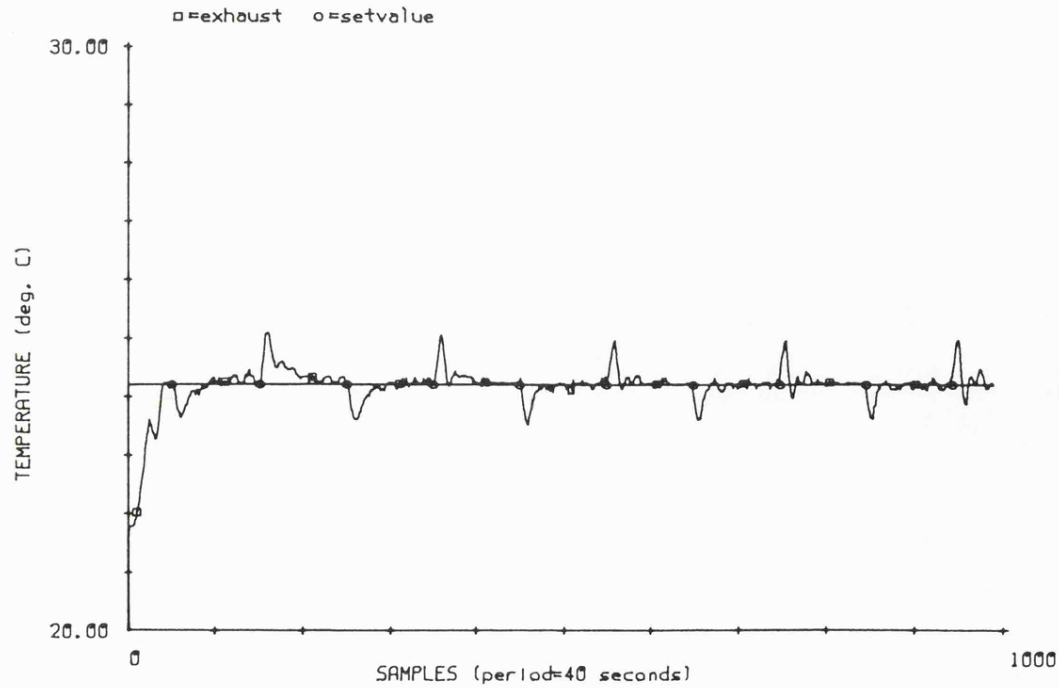


Fig. 5.2.18 Exhaust temperature response to radiant disturbance for conventional control and varying forward gain

Fig. 5.2.19 Exhaust temperature response to radiant disturbance for self-tuning control and varying forward gain



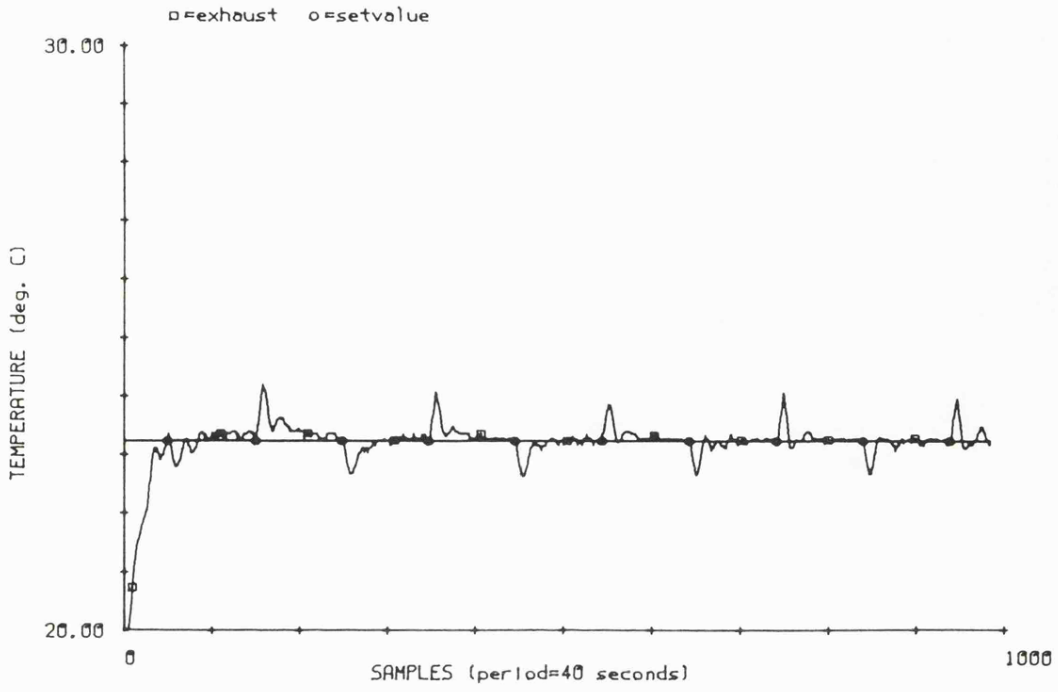
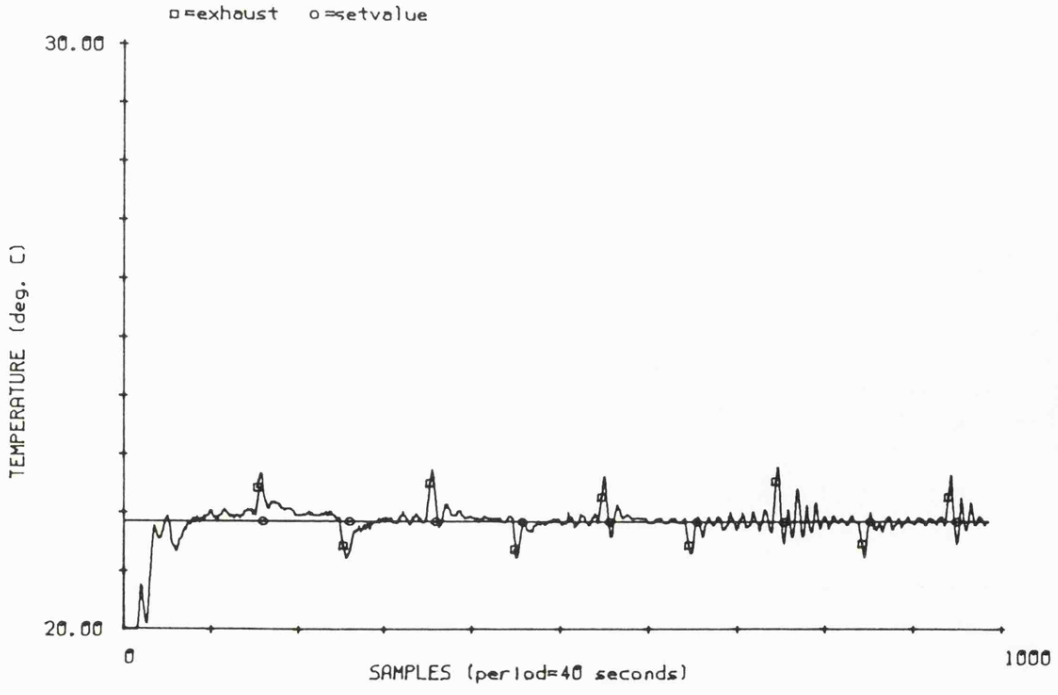


Fig. 5.2.20 Exhaust temperature response to radiant disturbance for self-tuning control with reduced parameters

Fig. 5.2.21 Exhaust temperature response to radiant disturbance for self-tuning control using inaccurate time delay estimate



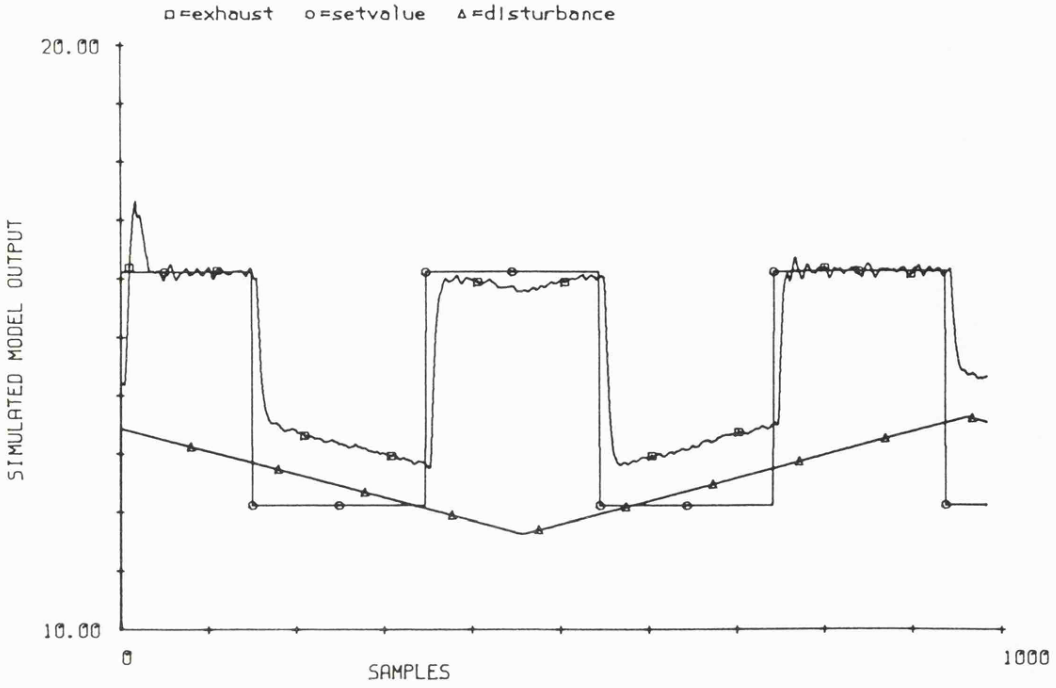
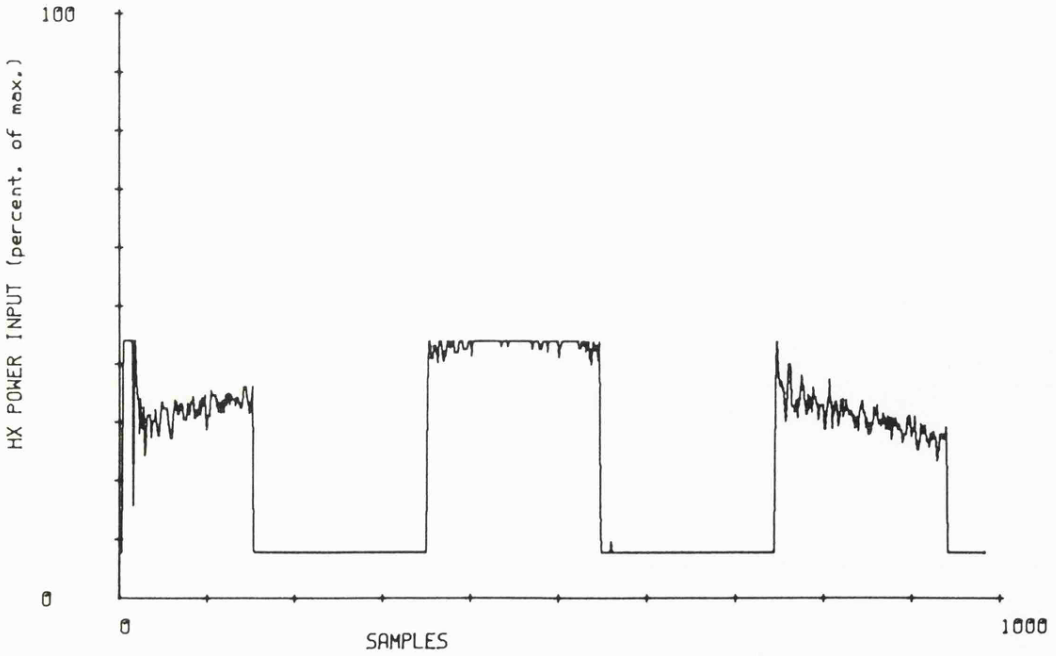


Fig. 6.1.1 Exhaust temperature response for self-tuning control , saturation limit reflected

Fig. 6.1.2 Heat exchanger input for self-tuning control



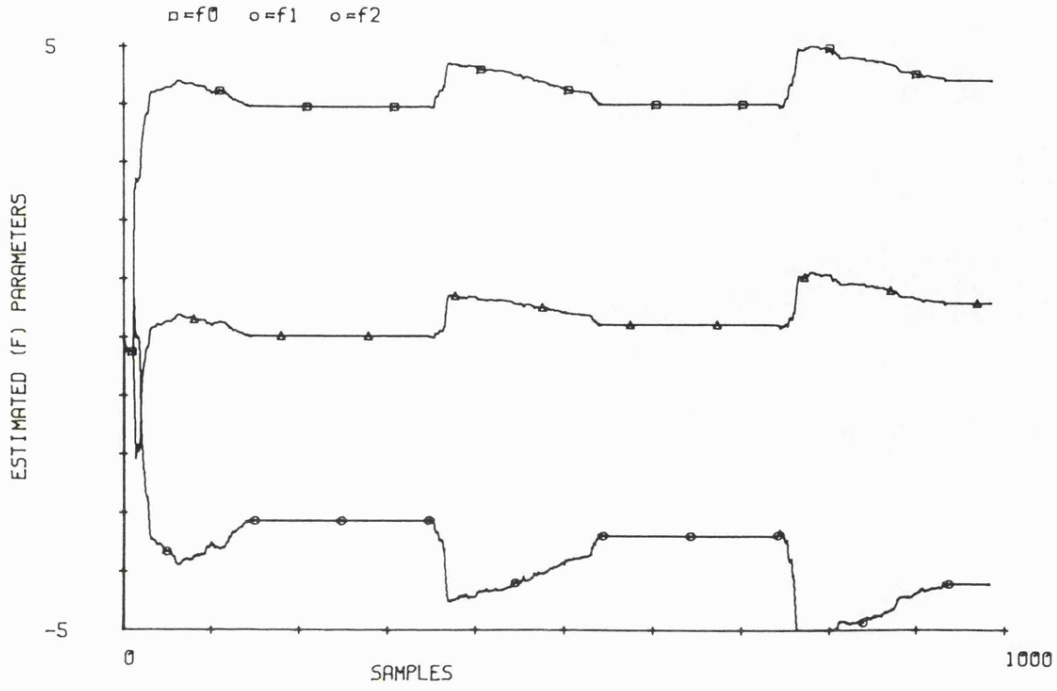


Fig. 6.1.3 'F' parameter estimates

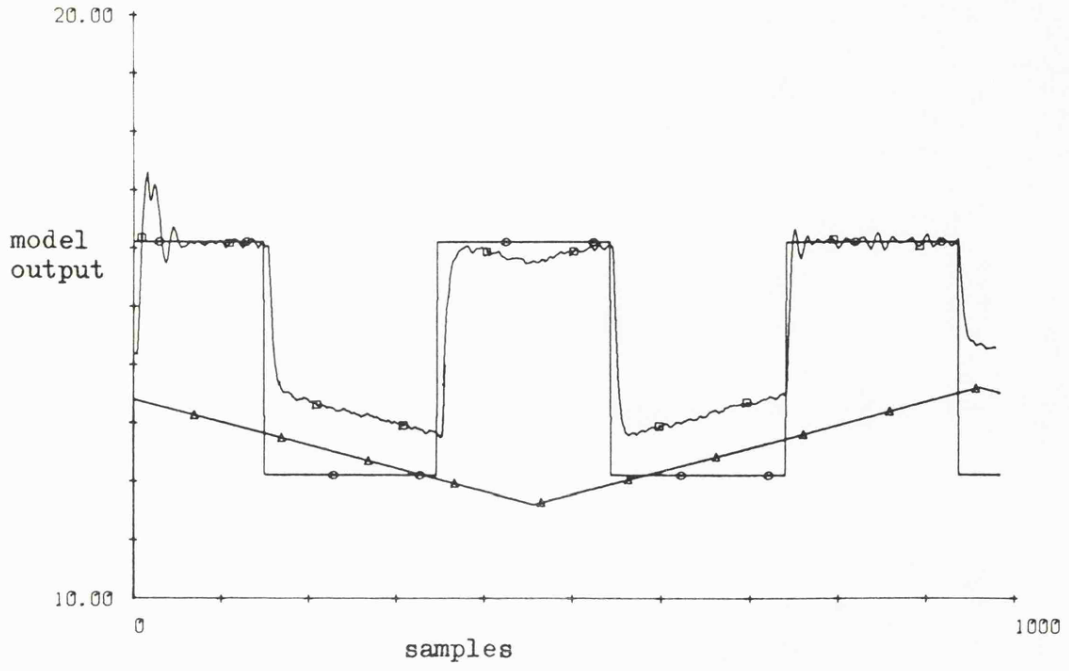
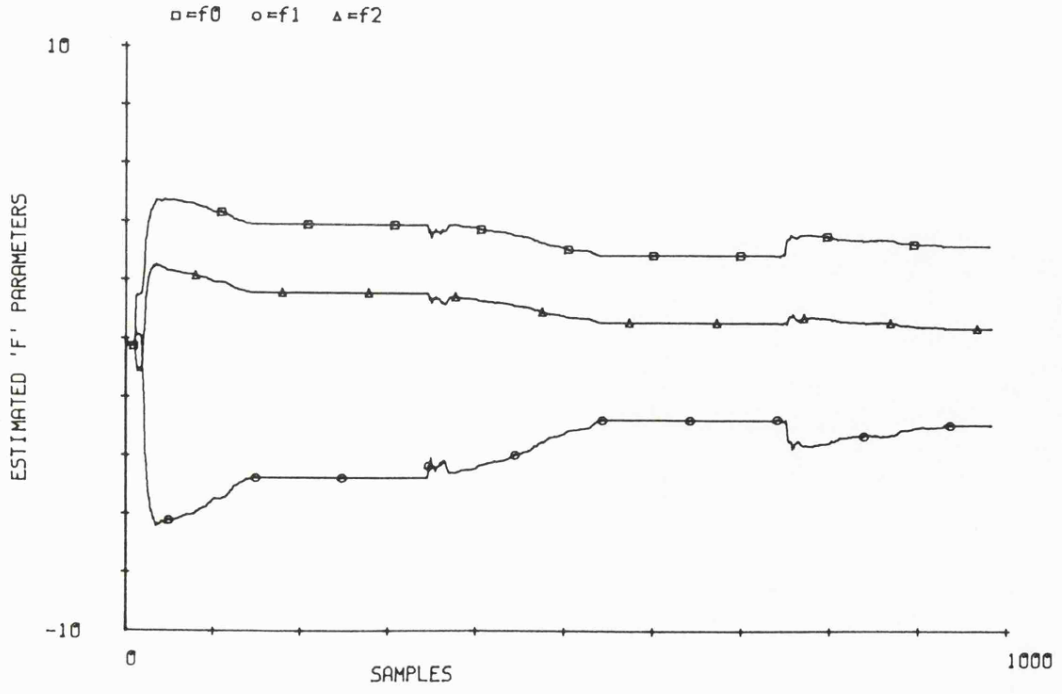


Fig. 6.1.4 Model output for self-tuning control , saturation limit NOT reflected

Fig. 6.1.5 'F' parameter estimates



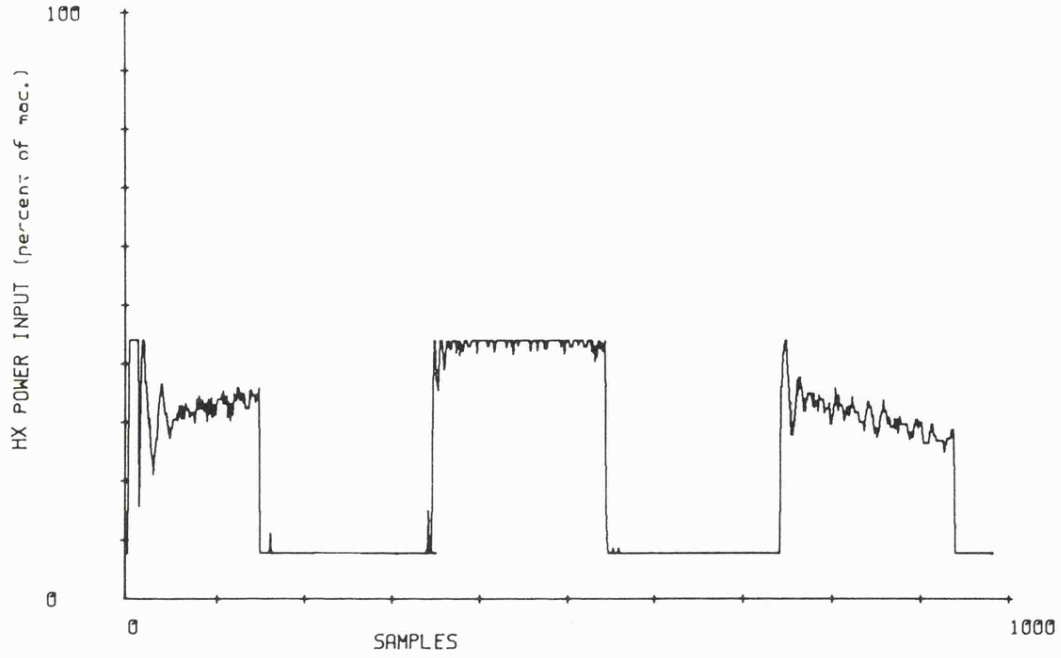


Fig. 6.1.6 Heat exchanger input for self-tuning control illustrating saturation

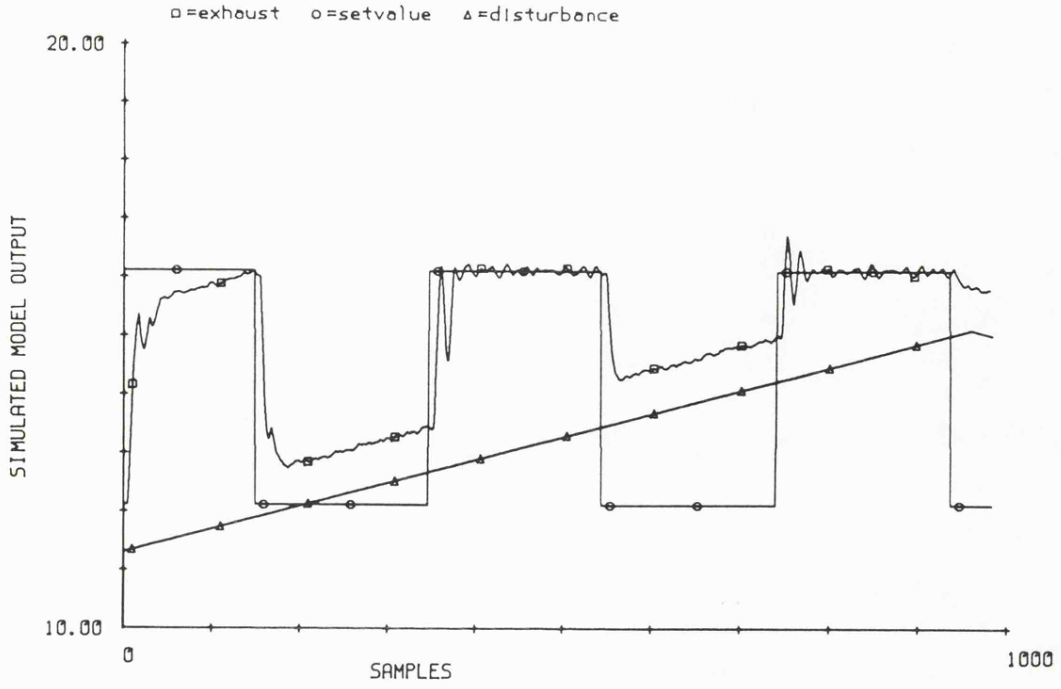


Fig. 6.1.7 Model output for self-tuning control , saturation ocurring during initial transient, limit reflected

Fig. 6.1.8 Model output for self-tuning control , saturation ocurring during initial transient , limit NOT reflected

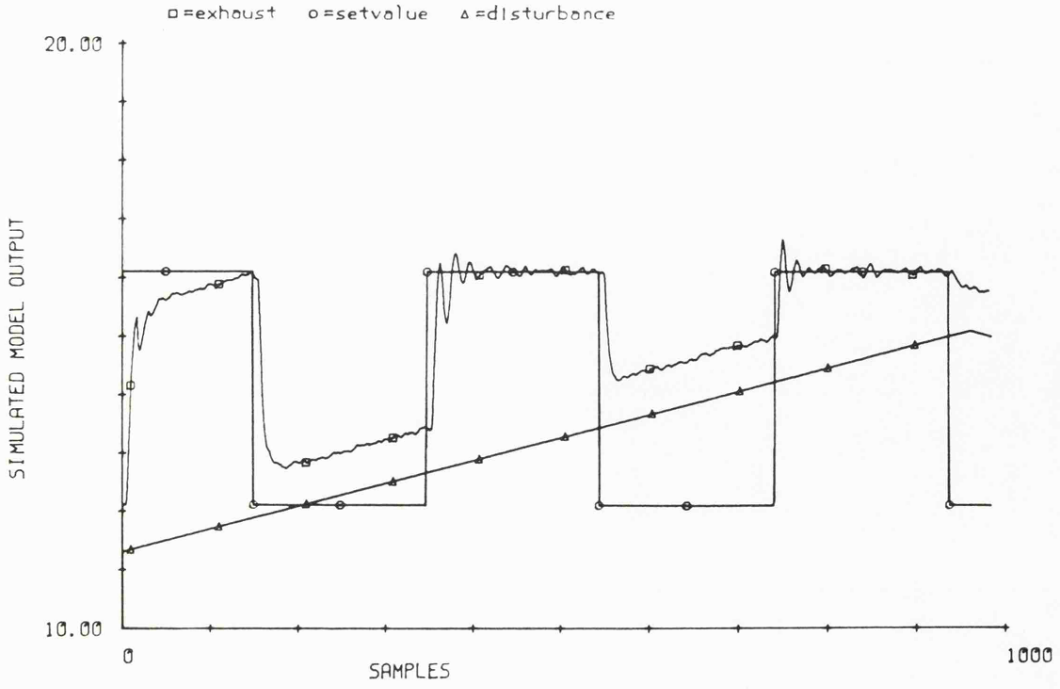
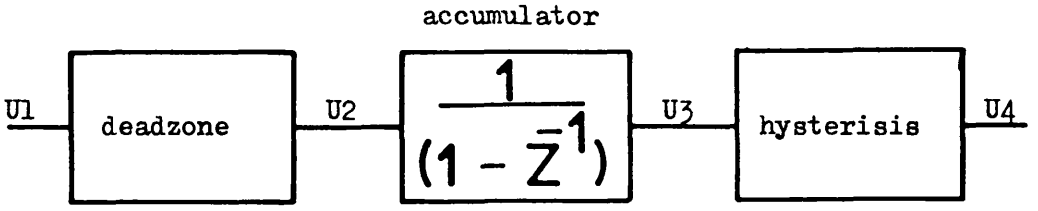


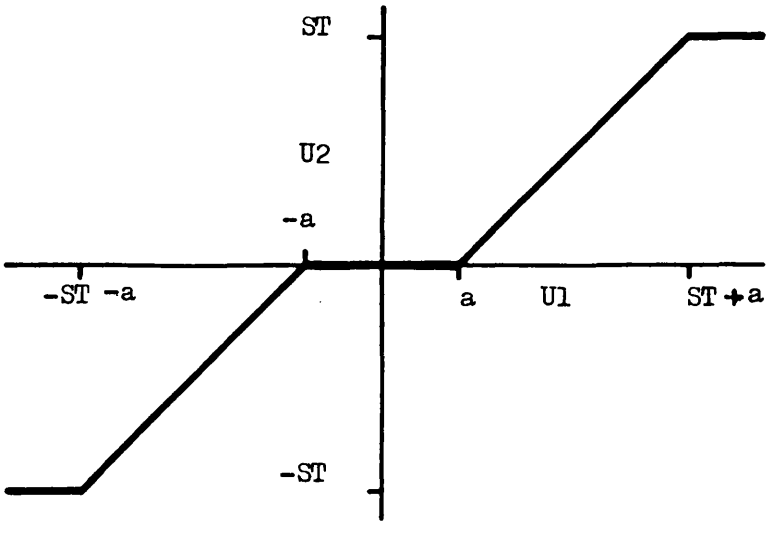
Fig. 6.2.1 Block diagram of simplified nonlinearities relating control signal and valve position



U1 = control signal

U4 = valve stroke

Fig. 6.2.2 Deadzone nonlinearity



ST = sample time

a = deadzone

Fig. 6.2.3 Hysterisis nonlinearity

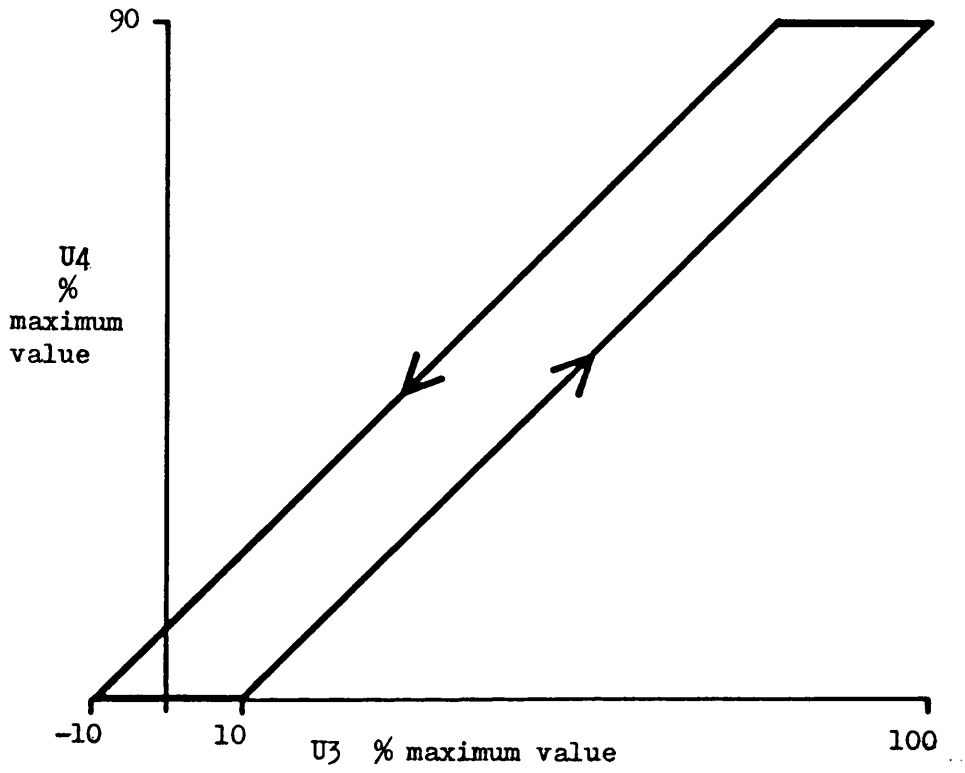


Fig. 6.2.4 Pseudo-code listing of anti-windup and deadzone compensation algorithm

```
;CONTROL...calculated control signal
;DEADZONE..symmetrical deadzone in seconds
;SAMPLE....sample time in seconds
;
COMPAIG:
DO
;accumulate control signal
       CONTROL = SUM + CONTROL
;if control signal is greater than sample time then apply limit
;save excess and return , otherwise continue
       IF  ABS( CONTROL ) >SAMPLE
       THEN
              SUM = CONTROL - SGN( CONTROL ) * SAMPLE
              CONTROL = SGN( CONTROL ) * SAMPLE
              RETURN
       ENDIF
;if control smaller than deadzone then save sum and set control
;to zero , and return , otherwise continue
       IF  ABS( CONTROL ) < DEADZONE
       THEN
              SUM = CONTROL
              CONTROL = 0
              RETURN
       ENDIF
;set sum to zero and return
       SUM = 0
       RETURN
END COMPAIG:
```

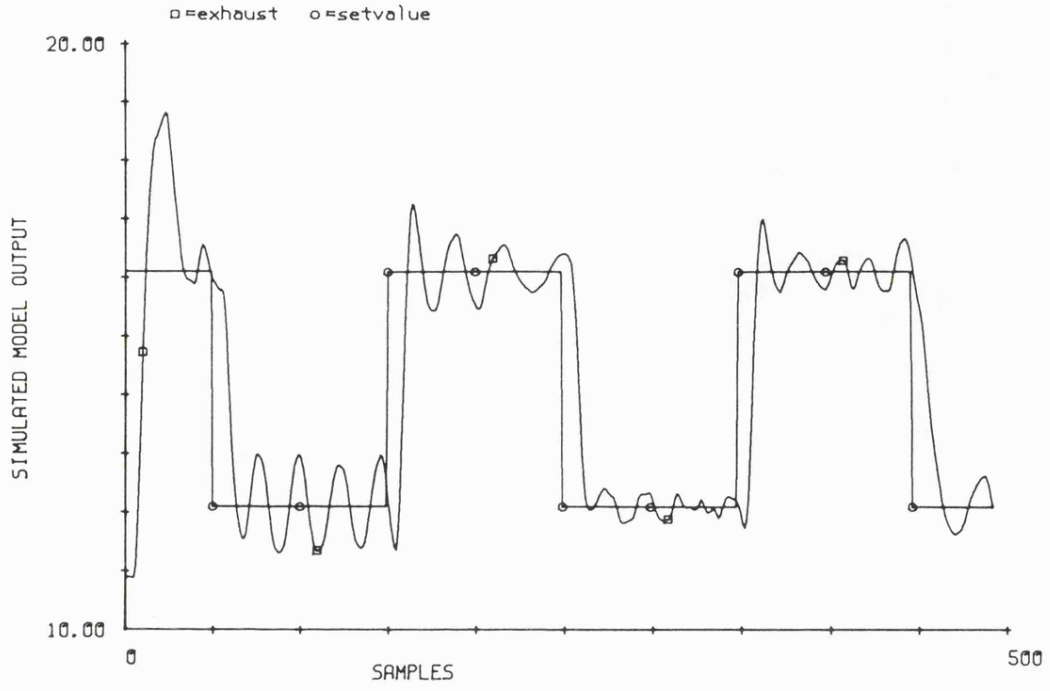
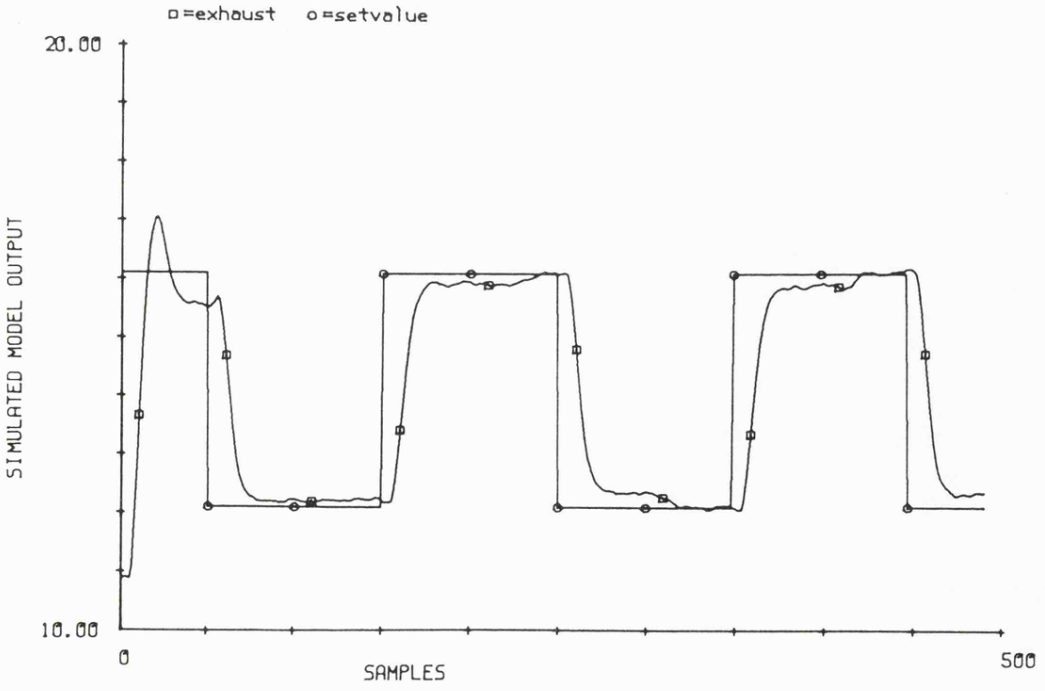


Fig. 6.2.5 Model output for self-tuning control and hysterisis equal to 10%

Fig. 6.2.6 Model output for conventional control and hysterisis equal to 10%



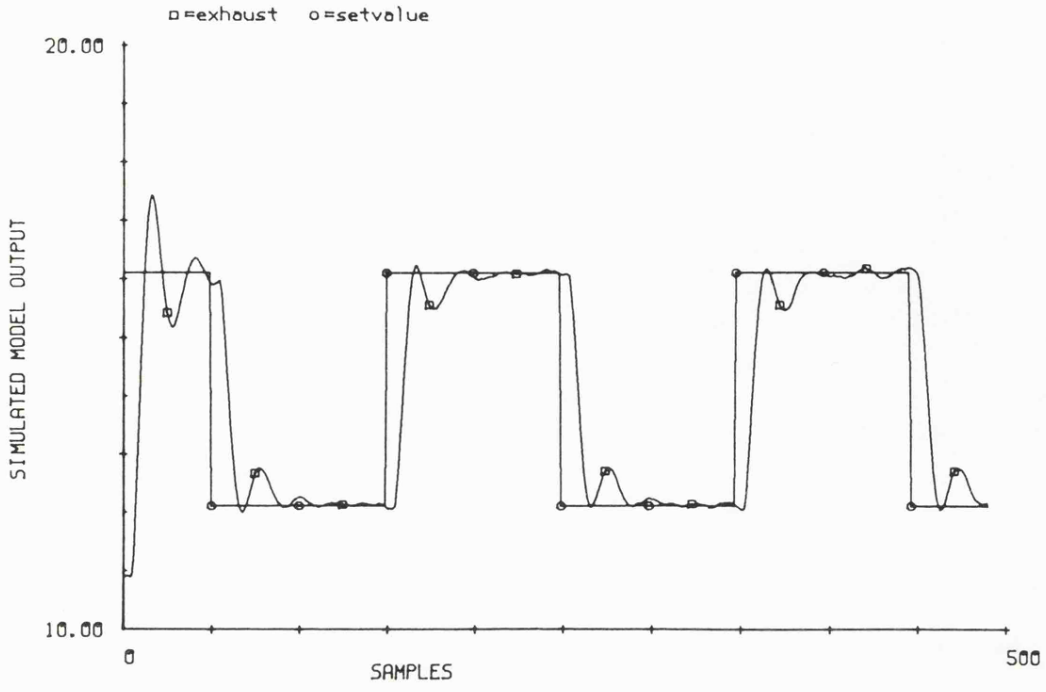
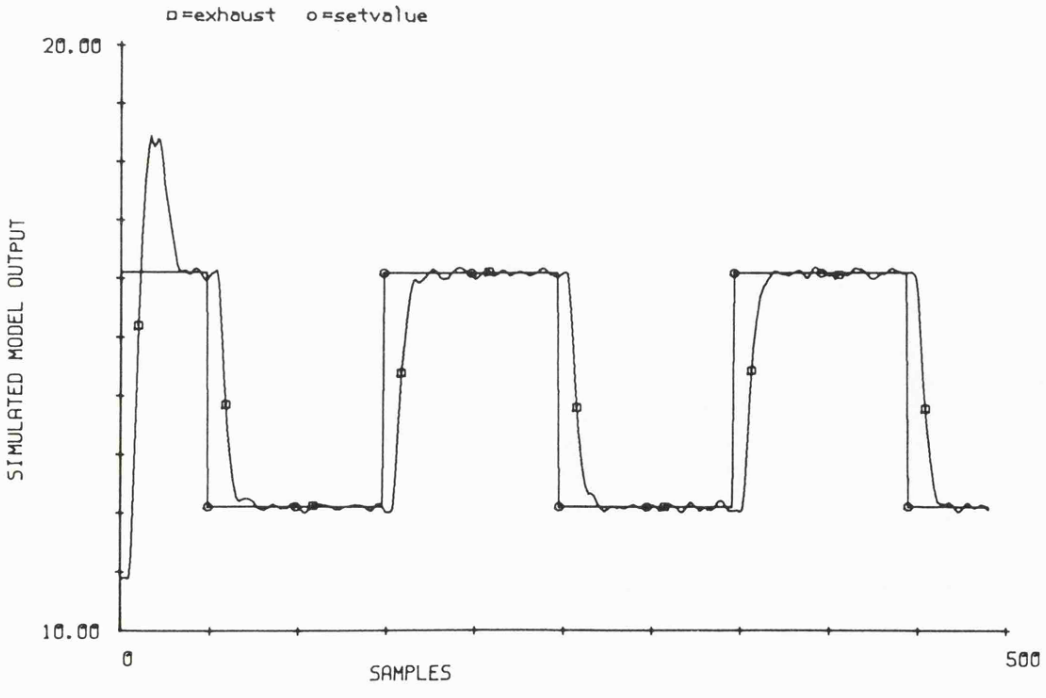


Fig. 6.2.7 Model output for conventional control and hysterisis equal to 10% , compensation equal to 10%

Fig. 6.2.8 Model output for self-tuning control and hysterisis equal to 10% , compensation equal to 10%



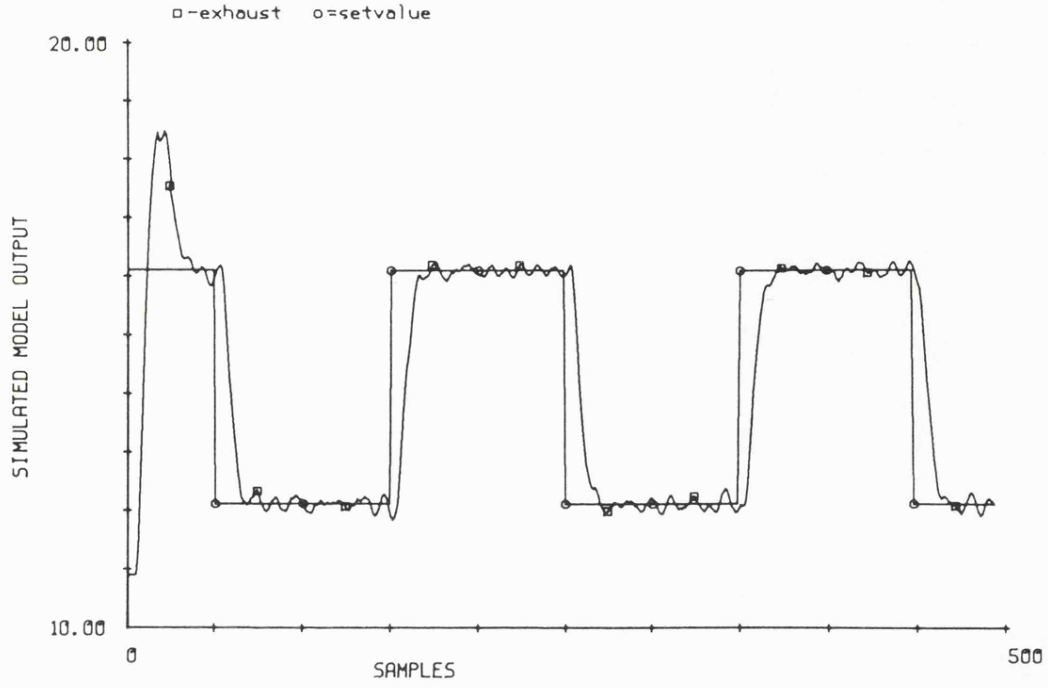


Fig. 6.2.9 Model output for self-tuning control and hysterisis equal to 10% , compensation equal to 12.5%

Fig. 6.2.10 Model output for self-tuning control and hysterisis equal to 10% , compensation equal to 7.5%